

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2003, 23:06:19 ; Search time 2628 seconds
(without alignments)
7408.591 Million cell updates/sec

Title: US-09-252-691c-1394

Perfect score: 669
Sequence: 1 gctaatgcgcgaattgat.....ttacccttaaggagaactga 669

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 20

Total number of hits satisfying chosen parameters: 20

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg_hum.*
- 31: em.htg_inv.*
- 32: em.htg_other.*
- 33: em.htg_mus.*
- 34: em.htg_pln.*
- 35: em.htg_rod.*
- 36: em.htg_mam.*
- 37: em.htg_vrt.*
- 38: em.sy.*
- 39: em.htgo_hum.*
- 40: em.htgo_mus.*
- 41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	23	3.4	10959	1 AE000213	AE000213 Escherich
C 2	23	3.4	11732	1 AE005329	AE005329 Escherich
C 3	23	3.4	15007	1 D90748	D90748 Escherichia
C 4	23	3.4	20284	1 D90749	D90749 Escherichia
C 5	23	3.4	110000	2 AC073744_0	AC073744 Mus muscu
C 6	23	3.4	222605	1 AP002555	AP002555 Escherich
C 7	22	3.3	24578	1 AE008754	AE008754 Salmonell
C 8	22	3.3	254050	1 AL627269	AL627269 Salmonell
C 9	21	3.1	10393	1 AE013675	AE013675 Versinia
C 10	21	3.1	216050	1 AJ414157	AJ414157 Versinia
C 11	20	3.0	884	8 AF415022	AF415022 Chamaebat
C 12	20	3.0	7770	1 AB064593	AB064593 Escherich
C 13	20	3.0	10473	1 AE013782	AE013782 Versinia
C 14	20	3.0	32403	3 U5370	U5370 Caenorhabdi
C 15	20	3.0	137586	9 AC106775	AC106775 Homo sapi
C 16	20	3.0	152037	2 AC128922	AC128922 Rattus no
C 17	20	3.0	171370	9 AC004021	AC004021 Human PAC
C 18	20	3.0	175838	2 AC116356	AC116356 Homo sapi
C 19	20	3.0	193050	1 AJ414149	AJ414149 Versinia
C 20	20	3.0	195829	9 AL353689	AL353689 Human DNA

ALIGNMENTS

RESULT 1
AE000213/c
LOCUS Escherichia coli K12 MG1655 section 103 of 400 of the complete genome. BCT 01-DEC-2000
DEFINITION Escherichia coli K12 MG1655 section 103 of 400 of the complete genome.
ACCESSION AE000213 U00096
VERSION AE000213.1 GI:1787371
KEYWORDS Escherichia coli K12.
SOURCE Escherichia coli K12.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 10959)
AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.
TITLE The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5331), 1453-1474 (1997)
MEDLINE 97426617
PUBMED 9278503
REFERENCE 2 (bases 1 to 10959)
AUTHORS Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE 3 (bases 1 to 10959)
AUTHORS Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE 4 (bases 1 to 10959)
AUTHORS Plunkett, G. III.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

COMMENT

This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, Director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES

source	1..10959	Location/Qualifiers
	/organism="Escherichia coli K12"	
	/strain="K12"	
	/sub_strain="MG1655"	
	/taxon:83333"	
	13..40	
promoter	/note="factor Sigma70; predicted +1 start at 1184959"	
protein_bind	119..136	
	/note="central position to predicted promoter:82"	
	/bound_molety="PhoB predicted site"	
gene	155..1381	
	/gene="pept"	
	/note="b1127"	
CDS	155..1381	
	/gene="pept"	
	/EC.number="3.4.11.-"	
	/function="putative enzyme; Not classified"	
	/note="o408; 100 pct identical to 42 aa fragment	
	PEPT_ECOLISW: P29745; 93 pct identical to 407 residues of	
	409 aa PEPT_SALTY SW: P26311"	
	/codon_start=1	
	/transl_table=11	
	/product="putative peptidase T"	
	/protein_id="AAC74211.1"	
	/db_xref="GI:1787372"	
	/translation="MDKLLERFLNYSLDTSQKAGVQVPSVTCQWKLHLLKEQLEE	
	MLGLNVLSKGLTALPLANPGDIPAIQFISHVDTSPDCSGKNVNPQIVENYRGD	
	IALGIDGIVSPVMPVFLHQLGLTITDGTGLGADDAKAGIAEINFTALAVLQQRKI	
	PHGDIRVAFPTDEVGKAGHFVDAPDAFWAYTVDGGVGELEFENFAASVNIKIV	
	GNNVHPTAGVMVNAISLAARTHAEVPADESPEMTGEYGEFVHLASMKCTVERADMH	
	YILRDFRQFOEARKRKMEIAKKVGLKLPDCYIELVIEDSYNNREKVVVEPHILD	
	IAQVIRAEALTAQRK"	
	complement(1430..2560)	
gene	/gene="ycfB"	
	/note="b1128"	
CDS	complement(1430..2560)	
	/gene="ycfB"	
	/function="orf; Unknown"	
	/note="f376; 100 pct identical to fragment YCFD_ECOLI	
	SW:P27431 but has 3 additional N-terminal aa and 93	
	additional C-terminal aa"	
	/codon_start=1	
	/transl_table=11	
	/product="orf, hypothetical protein"	

/protein_id="AAC74212.1"	
/db_xref="GI:1787373"	
/translation="MLNMEYQLTLLNWPDFLERHWKRPVVLKRGFNFIIDPISPDELA	
GLAMESRSLYSHQDGKQVSHGSPESDYHLGETNWSLLVQAVNHHWHEPHTALMRP	
FRELDPNRIDDLMSFSPVGGVGHLDQVDVFIQTGRRRVGKGLQMKHCCHP	
DLQVDFEALIDEELEPGDILYIPPGPHGEGYALENAMNYSVGFRAFPNRELISGFA	
DYVLQRELGNYSDDPPRAHPADVLQPMKLEMLLELINQPHHFKQWFEFTIS	
QSRHLEDAIEPPYQDEIYDALKQCEVLVRLGGLRVLRIGDDVYANGEKIDSPHRP	
ALDASNIALTAEFGDALEDPSFLAMLALVNSGVWFEG"	
complement(2583..2613)	
/note="factor Sigma70; predicted +1 start at 1187488"	
complement(2627..4087)	
/gene="phoQ"	
/note="b1129"	
complement(2627..4087)	
/gene="phoQ"	
/EC.number="2.7.3.-"	
/function="enzyme; Global regulatory functions"	
/note="f486; 99 pct identical to PHOQ_ECOLI SW: P23837"	
/codon_start=1	
/transl_table=11	
/product="sensor protein PhoQ"	
/protein_id="AAC74213.1"	
/db_xref="GI:1787374"	
/translation="MKLLRFFLSLRVRLATAAVLVLSLAYGMVALIGYSVSF	
DKTTFLLRGESNLFYTLAKWNNKHLVELPENIDKQSTMTLIYDENGQLLWQARDV	
PWLKMTQDPLKSNGFHETADVNDTSLLSGDHSTQQOQLQEVREDDDAEATHSVA	
VNYPATSRMPKLTIVVDPIPVLEKSSYMWKFIYVLSANILLVTPLLVAHWNSL	
RPIEALKEVRELEHNRLLNLTRELHSLVRLNRLKLSEREDYDKTTRTTLDT	
HSILAKLAVLQSTLSLRSEKMSVSDAEVPMLEQISRSQQOIGYILHRASMRGTLIS	
RELHPVAPLDLNTSALKVYQRKVNISLSDISPEISFVGQNDFVEVMGNVLDNACK	
YCLEFVEISARQTDHLYIVVEDDGGPILSKREVIFDRGQRVDTLRPGQGVGLAVAR	
EITEVEGKIVAGESMLGARMVEIFGRQHSAPKE"	
complement(4087..4758)	
/gene="phoP"	
/note="b1130"	
complement(4087..4758)	
/gene="phoP"	
/function="regulator; Global regulatory functions"	
/note="f223; 100 pct identical to PHOP_ECOLI SW: P23836"	
/codon_start=1	
/transl_table=11	
/product="transcriptional regulatory protein"	
/protein_id="AAC74214.1"	
/db_xref="GI:1787375"	
/translation="MRVLVVEDNALLRHHLKVLQDAGHOVDDAEADAEADYVLNEHI	
PDIAIVDLGLPDEGLSLIRWRSDVSLPILVLTAPESQDKVLEVSACADYVTKP	
FHIEEVMARQALMRNSGLASQVISLPFPQVLSRRELSINDEVIKLTAFETIMET	
LIRNNGKVSKDSLMQLQYLPDAELRESHTIDVLMGLURKKIQAQYPOEVITTVRGQSY	
LFEL"	
complement(4257..4287)	
/gene="phoP"	
/note="factor Sigma70; predicted +1 start at 1189162"	
complement(4297..6297)	
/gene="purB"	
/note="b1131"	
complement(4297..6297)	
/gene="purB"	
/EC.number="4.3.2.2"	
/function="enzyme; Purine ribonucleotide biosynthesis"	
/note="f456; 99 pct identical to PURB_ECOLI SW: P25739"	
/codon_start=1	
/transl_table=11	
/product="adenylosuccinate lyase"	
/protein_id="AAC74215.1"	
/db_xref="GI:1787376"	
/translation="MELSSLTAVSPVDGRYGDVKSALRGIFSEYGLLKFVQVEVRWL	
OKLAHAAIKEVPAPADAIGYDAIVASPSDEADAIRKTIERTNHDVRAVEFLKE	
KVAIEPDLHAVSEFIHFACSTEDINNLHSLMLKLTARDEVILPWRQLIDGINKDLAQ	
YRIPIPLSRTHGQAPATPTIGKEMANVAYRMERQYROLNQVEILKRGINGAVGNNAH	
ADYPEVDVHQFSEEFVTSLGIQWNPTTQLEPHDYIAELFDCEVAFKLTIDFDROVW	
GYIALNHFKOKTITAGEISMTMPKVPIDFENSEGNLSLNSLAVKSLKLPVSRWQ	
RDLTDSVLRNLGVGIGYALIAVQSTLKGVSKELVNRDHLDELHNWEVLAEPITQV	

```

gene      MRRYIEKPYEKLKELTRGKRVDAEGMKQFDGLALPEEEKARLKAMTPANYIGRAIT
          MVDELK"
          complement(6301. .6942)
          /gene="ycfc"
          /note="b1132"
          complement(6301. .6942)
          /gene="ycfc"
          /functions="orf; Unknown"
          /note="f213; 99 pct identical to YCFC_ECOLI SW: P25746"
          /codon_start=1
          /transl_table=11
          /product="orf, hypothetical protein"
          /protein_id="AAC74216.1"
          /db_xref="GI:1178737"
          /translation="MAKNYYDITLALAGICQSRARLVQOLAHQGHCDADALHVSLSNII
          DMPSTLAVFGGSEANLRVGLTGLVNLASSRQGLNAELTRYTLISLMVLERKLSSA
          3.4% Score 23; DB 1; Length 10959;
          Best Local Similarity 100.0%; Pred. No. 0.19;
          Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
          Qy 133 CTGTTCAATAAACCTAGCATGT 155
          |||||
          Db 9136 CTGTTCAATAAACCTAGCATGT -9114
          |||||
          Query Match
          3.4% Score 23; DB 1; Length 10959;
          Best Local Similarity 100.0%; Pred. No. 0.19;
          Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
          Qy 133 CTGTTCAATAAACCTAGCATGT 155
          |||||
          Db 9136 CTGTTCAATAAACCTAGCATGT -9114
          |||||
```

```

RESULT 2
AE005329/c 11732 bp DNA linear BCT 21-MAR-2001
LOCUS Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 153
of 155.
ACCESSION AE005329
VERSION AE005329.1 GI:12514784
KEYWORDS Escherichia coli O157:H7 EDL933.
SOURCE Escherichia coli O157:H7 EDL933
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 11732)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouisis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)
21074935
PUBMED 11206551
REFERENCE 2 (bases 1 to 11732)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouisis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Direct Submission
Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
LOCATION/Qualifiers
1. .11732
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohemorrhagic"
complement(115. .1485)
/gene="purB"
/note="Z1860"
complement(115. .1485)
/gene="purB"
/functions="enzyme; Nucleotide biosynthesis; Purine
ribonucleotide biosynthesis"
```

```

/note="Residues 1 to 456 of 456 are 99.78 pct identical to
residues 1 to 456 of 456 from Escherichia coli K-12 Strain
MG1655: B1131"
/codon_start=1
/transl_table=11
/product="adenylosuccinate lyase"
/protein_id="AAG55957.1"
/db_xref="GI:12514785"
/translation="MELSLTAVSPVDGRYGVKVSALRGIFSEYGLLKFRVQVEVRWL
OKIAHAHAIKEVFAFAADAIGYDAIVANSEEDAAIKTERTTNDHVKAVEYFLKE
KVAIEPILHAVSEFIFHACTSEIDNNLSHMLKRTARDEVILPYWRQIDIGKDLAYQ
YRDIPLLSRTHGQPATPSTIGKMANVAYRMERYROLNOVEILGKINGAVGNNAHI
ADPEVDWHQFSEEFVTSIGKONNPTTQIEPHDYIAELFDCVARFNTILIDEDRDYV
GYALNHFKAKTTAGEIGSSTMPHKYNPIDFENSEGNLGSNAVLQHLAKLPSVRNQ
RDTYDSTVLNKLGVIGYALGYAIDGKLVSKLEVRNDRDLDELHNEVLAEPITQV
MRRYIEKPYEKLKELTRGKRVDAEGMKQFDGLALPEEEKARLKAMTPANYIGRAIT
MVDELK"
complement(1489. .2130)
/gene="ycfc"
/note="Z1861"
complement(1489. .2130)
/gene="ycfc"
/functions="orf; Unknown function"
/note="Residues 1 to 213 of 213 are 99.53 pct identical to
residues 1 to 213 of 213 from Escherichia coli K-12 Strain
MG1655: B1132"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55958.1"
/db_xref="GI:12514786"
/translation="MAKNYYDITLALAGICQSRARLVQOLAHQGHCDADALHVSLSNII
DMPSTLAVFGGSEANLRVGLTGLVNLASSRQGLNAELTRYTLISLMVLERKLSSA
KGALDTLGNRLNGIQLEHFDLQSLMSAAIYVDVISPGLQVIGTQVGGSPAVLOS
POQAKYVRATLLAGIRAAVLVHQVGGGRLQLMFSRNLRTTQAKQILAHLPTEL"
complement(2166. .3317)
/gene="ycfb"
/note="Z1862"
complement(2166. .3317)
/gene="ycfb"
/functions="orf; Unknown function"
/note="Residues 1 to 383 of 383 are 99.47 pct identical to
residues 1 to 383 of 383 from Escherichia coli K-12 Strain
MG1655: B1133"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55959.1"
/db_xref="GI:12514787"
/translation="MTEYAALKENYVSDPMSETAKKVIYVGMGGVSDSVSNWLLQOQG
YQVEGLFMKNWEEDGEEYCTAADLADAQVCDKLGIELHTVNFAYEDWNVFELFL
AEYKAGTPNPDIKCNKEIKFAPLEAEDLADYIATGHYVRADVDGKSRLLRGL
DSNRQSYFLYLSHQEIAQSLFPGVELEKQVQRKIAEDLGLTAKKDGTCGFCGE
RKRFELGYLPAQPGKIIITVDGDEIGEHQGLMYHTLQGRKGLGIGTKGTEPWT
TKVDLVNNILIVAGQHEHPLMSVGLIAQOLHWVDREPFTGTRCTVKTVRTQTDIPC
TVKALDDRIEIVFDEPVAATPGCSAVFYNGEVLGGGIIIEQRLPI"
complement(3326. .3787)
/gene="Z1863"
complement(3326. .3787)
/gene="Z1863"
/functions="putative enzyme; Not classified"
/note="Residues 1 to 153 of 153 are 100.00 pct identical
to residues 1 to 153 of 153 from Escherichia coli K-12
Strain MG1655: B1134"
/codon_start=1
/transl_table=11
/product="putative phosphohydrolase"
/protein_id="AAG55960.1"
/db_xref="GI:12514788"
/translation="MFKPHVTACVHHAEGKFLVVEETINGKALWNOPAGHLEADEFL
VEAAARELWEETGISAQPHFIRVMHQAIPDKTFLFLFAIELEICQICPQHDSID
CCRWVSAEIILOASNLRLSPLVAESIRCYQSGORYPLEMIGDFNWPFTKGI"
complement(3797. .4420)
```

```

/ gene="ymfc"
/ note="Z1864"
/ complement(3797..4420)
/ gene="ymfc"
/ function="orf; Unknown function"
/ notes="Residues 1 to 207 of 207 are 99.03 pct identical to
residues 1 to 207 of 207 from Escherichia coli K-12 Strain
MG1655; B1136"
/ codon_start=1
/ transl_table=11
/ product="orf, hypothetical protein"
/ protein_id="AAG55961.1"
/ db_xref="GI:12514789"
/ translation="MQKTSFRNQVRFSSQSTRRKPNQPTRVILFNKPYDVLQPF
TDAGAKTKLEFIPQGVGAAGRLDRDSEGLVLTNNGALQARLTQPGKRTGKIYVQ
VEGIPQDALEALRNGVLTNDGPTLPAGAEVDEPAWLWPNPPIREKRSIPTSLKI
TLVEGRNROVRMTAHVGPFTLLIRIYANGDYSLDNLANGWEARE"
4622..5872
/ gene="icdA"
/ note="Z1865"
4622..5872
/ gene="icdA"
/ function="enzyme; Energy metabolism, carbon: TCA cycle"
/ notes="Residues 1 to 416 of 416 are 99.27 pct identical to
residues 1 to 416 of 416 from Escherichia coli K-12 Strain
MG1655; B1136"
/ codon_start=1
/ transl_table=11
/ product="isocitrate dehydrogenase, specific for NADP+"
/ protein_id="AAG55962.1"
/ db_xref="GI:12514790"
/ translation="MKSIVVPAQKKITLQNKLVNPNPIPYIEGGIGVDVTPA
MLKVDAAGVAKYKGRKISWMEIYTGKSTQYGVQVLPAPETLRLIREYVAIRGP
LTPVGGIRSLNVAQLQELDYICLRPVYQYQVSPKHPBELTDWVIFRENSDIY
AGTWDVADAKVTKFLREMGVKKIRFPEHCIGIKPCSEGTKLVRALAEVAI
ANDRSVTLVHKINIMKTEGAFKDWGLYELAFEGGELIDGGPMLKVNPTGKEIV
IKDVIADAFLOQLLRPAEDYVACNNGDYISDALAAQVGGIGTAPANGTDECAL
FEATHGATKACQVNFSGSIILSAEMMLRMGMWTEADLIYKMGEGAINAKTVTD
FERLMGALKKCGSEFGAIIKNN"
5791..>11732
/ note="O-island #52; Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655"
/ complement(5986..7128)
/ gene="Z1866"
/ complement(5986..7128)
/ gene="Z1866"
/ function="putative enzyme; Integration, recombination
(Phage or Prophage Related)"
/ notes="Residues 1 to 380 of 380 are 98.42 pct identical to
residues 1 to 380 of 380 from Genpept 118 :
9112154501gb|AAA32336.1| (M61865) integrase [Bacteriophage
21]"
/ codon_start=1
/ transl_table=11
/ product="putative integrase of prophage CP-933X"
/ protein_id="AAG55963.1"
/ db_xref="GI:12514791"
/ translation="MAASPRSHKISIPNLYCKLDRKTKGVYQYKHPLSGRFSLGTD
ENKAKVATEANTIAEQTRTOILSVNERLKMGRSDITVTEWLDKYNISQEDRLQ
HNELRNSYRQKGPRLRFEHCGMOHLKDITADIAELIDAVKAEHNRMAQVPMV
LIDVFEAGHAGVHPGFPNPAQATKQPRNVNRQRLSFLWQAFISVSRQPYLKG
MLCALVYQRLSDICNLKFSIDWDDMLHITQETKSKLAIPNLKCALNITLREVIS
QCDVAVSYLYHYRHTTSQANRGDQVSANTLTAFKAKREKCGIKWEGQTAPTFEQ
RSLSLRYEQGLDTOKLLGHKSRKMTDRYNDRDKDIIVDIKTA"
/ complement(7118..7354)
/ gene="Z1867"
/ complement(7118..7354)
/ gene="Z1867"
/ function="putative enzyme; Integration, recombination
(Phage or Prophage Related)"
/ notes="Residues 1 to 78 of 78 are 98.71 pct identical to
residues 1 to 78 of 78 from Genpept 118 :
9112154491gb|AAA32335.1| (M61865) excisionase

```

```

[ Bacteriophage 21 ]
/codon_start=1
/transl_table=11
/product="putative integrase of prophage CP-933X"
/protein_id="AAG55964.1"
/db_xref="GI:12514792"
/translation="MSRLTILRDWAKEEFGDLAPSERVLKVAQGMMAIPAIVGRY
WMIDRNSREVGTGLAEPLPINAPKLQRIADGC"
7458..8201

```

```

gene
Query Match 3.4%; Score 23; DB 1; Length 11732;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 133 CTGTTCAATAAACCCCTACGATGT 155
|||||
DB 4324 CTGTTCAATAAACCCCTACGATGT 4302
|||||

```

```

RESULT 3
D90748/c
LOCUS D90748 15007 bp DNA linear BCT 07-FEB-1999
DEFINITION Escherichia coli genomic DNA. (25.6 - 25.9 min).
ACCESSION D90748 AB001340
VERSION D90748.1 GI:1651553
KEYWORDS Complete and shotgun sequencing; potB; potA; pepT; phoQ; phoP;
purB; ycfC; ycfB; icdA; icd; icdE; lit.
SOURCE Escherichia coli(strain:K12) DNA, clone:Kohara clone #239.
ORGANISM Escherichia.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (sites)
REFERENCE
AUTHORS Oshima, T., Itoh, T., Kajihara, M., Kanai, K., Kashimoto, K.,
Kishimoto, K., Inada, T., Itoh, T., Kimura, S., Miki, T.,
Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishimoto, H.,
Nishio, Y., Saito, N., Sampei, G., Seki, Y., Tagami, H., Takemoto, K.,
Wada, C., Yamamoto, Y., Yano, M. and Horiuchi, T.
A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map
DNA Res. 3 (3), 137-155 (1996)
97061202
2 (sites)
REFERENCE
AUTHORS Alba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A.,
Horiuchi, T., Isono, K., Itoh, T., Kanai, K.,
Kashimoto, K., Kim, S., Kimura, S., Kitagawa, M.,
Kitakawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H.,
Motomura, K., Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T.,
Saito, N., Sampei, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,
Yamamoto, Y. and Yano, M.
The systematic sequencing of the Escherichia coli genome in Japan
Unpublished
REFERENCE
AUTHORS Mori, H.
3 (bases 1 to 15007)
Direct Submission
Submitted (29-JUL-1996) Hirotsada Mori, NARA Institute of Science
and Technology, Res. & Edu. Center for Genetic Info.: 8916-5
Takayama, Ikoma, Nara 630-01, Japan
(E-mail:hmori@etci.aist-nara.ac.jp, Tel:81-7437-2-5660,
Fax:81-7437-2-5669)
Collaboration Information:
Project:
The Japan E.coli genome DNA sequencing project
Group:
The Japan E.coli genome DNA sequencing group
Members: (1995.4 - 1996.3)
Alba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A.,
Horiuchi, T., Ikemoto, K., Inada, T., Isono, K., Isono, S.,
Itoh, T., Kanai, K., Kasai, H., Kashimoto, K., Kim, S.,
Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K.,
Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K.,
Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T., Saito, N.,
Sampei, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,

```

Yamamoto, Y. and Yano, M.

Headed by:

Name: Takashi Horiuchi

Address: National Institute of Basic Biology, Okazaki, 444, Japan

E-mail: kishori@nibb.ac.jp

Information operator:

Name: Hirotada Mori

Address: NARA Institute of Science and Technology,

Ikoma, 630-01, Japan

E-mail: hmori@gtc.aist-nara.ac.jp

URL:

The Japan E. coli genome database

http://bsw3.aist-nara.ac.jp.

Location/Qualifiers

1..15007

/organism="Escherichia coli"

/strain="K12"

/db_xref="taxon:562"

/map="25.6-25.9 min"

/clone="Kohara clone #239"

/note="Nucleotide position 1185468-1200474 from the initiation site of *thiA* (0 min.). This clone is from Kohara lambda miniset library."

complement(1..678)

/gene="potB"

complement(<1..678)

/gene="potB"

/note="ORF_ID:0238#14"

similar to SwissProt Accession Number P23860"

/codon_start=1

/transl_table=11

/product="Spermidine/putrescine transport system permease

protein PotB."

/protein_id="BAA35947.1"

/db_xref="GI:1651554"

/translation="MIVTIVGWLVLVFLPLNMIIGTSFLTRDDASFVKMVFLLDNYT RLDDPLFEVLHLSNLMALIALACVLVGFPAFWLAKLPHKVRFLLLVLPFWTN SLIRTYGLKIFLSYKGLNELLWGLVDTPTRIMFTPSAVTIGLVILLPFWMPLY SSIEKDKPLLEAARDLGASKLQTFIRIILPTMPTIAGCLLVMLPAMGLFYVSDLM GGAKNLI"

complement(692..1828)

/gene="potA"

complement(692..1828)

/gene="potA"

/note="ORF_ID:0239#1"

similar to PIR Accession Number A40840"

/codon_start=1

/transl_table=11

/product="Spermidine/putrescine transport protein A"

/protein_id="BAA35948.1"

/db_xref="GI:1651555"

/translation="MGQSKLNKQPSLSPLVLQAGIRKCFDGEVTPQLDLTINNGE FLTLGPSGCKTVLRLIAGLETVDSDGRIMLDNEDITHVPAENRYVNTVFSYALFP HMTVFAGFLRWQKTPAAETPRVMEALRMVQLETQAQRHQLSGGQQQQRVAIAR AVANKPRLLLDSELDYKLRQMQLKQKLGITFVFVTHDQEALTWSDRIC VVMRGRIEODGTREIYEKPLNFVAGFGEIINNFNATVIERDEQVRANVEGREG NITYNFAVEPGKHLVLRLEDLVEEINDONHAEGLLGYVRENYKGMTLESVVELE NGKMWVSEFNEDDPDFDHSLDQKMAINWVESVEVLADEEHK"

2078..3304

/gene="pept"

2078..3304

/gene="pept"

/note="ORF_ID:0239#2"

similar to SwissProt Accession Number P29745"

/codon_start=1

/transl_table=11

/product="Peptidase T (EC 3.4.11.-) (aminotripeptidase)

(tripeptidase) "

/protein_id="BAA35949.1"

/db_xref="GI:1651556"

/translation="MDKLLRFLNYSLDTSKAGVRQVPSTGEQKLLHLKLEQLEE MGLNVLSEKGTLMATLPANVPGDIPAIGTISHVDTPSDCSGKNVQIENVTRGGD IALGTGDEVLSPVMPFVLLHQLGTLITTDGKTLGADDDKAGIAEIMTALAVLQKKI

PHGDIRVAFTPDEEVGKGAKHFDVDAFDARWAYTVDGGGVGELEFENFNAASVNIKIV GNNVHPTAGKGVWVNLASLAARIHAIEVPADESPENTEGEYGHLSAMKGTVERADMH YIIIRDFRQKFARRKRMKEIAKKVKGKGLHPDCYIELVIEDYSIYNNRMEKVVHPHILD IAAQAMRDCDIEPELKPINGGTGAQLSFMGLPCPNLFTGGYNYHGKHBEFVTLGMEK AVQVIVRIAELTAQRK"

complement(join(3353..4474,4550..6010))

/gene="phoQ"

complement(3353..4474)

/gene="phoQ"

/note="ORF_ID:0239#3"

similar to PIR Accession Number D41966"

/codon_start=1

/transl_table=11

/product="Orf 2 downstream of phoQ."

/protein_id="BAA35950.1"

/db_xref="GI:4062695"

/translation="MEYQLTLNMPDFLERHWKRPVYLKRGNNFTIDPISPDLAGLA MESEVDSRLVSHQDQKQVSHGPFESYDHLGETNMSLLVQAVNHHHEPTAALMRPFRE LPDMRIDDLMIISFSPVGGVGHLDQYDFVFIQTGRRRWVGEKLMQKRELPHPDLL QVDPFEAIIIDELEPGDILYIPPGFPHGEGYALENAMNYSVGFRAPTRRELISGFADYV IQRELGGNYSDPDVPPRAHPADVLQPMDKLREMMLELINOPEHFKWQFGEFISQSR HELDITAPPEPPQPDIEYDALQGEVLYRGLGRVLRIGDDVIYANGKIDSPHRPALL ALASIALTAENFGDALEDPSEFLAALVNSGYWFFEG"

complement(4550..6010)

/gene="phoQ"

/note="ORF_ID:0239#4"

similar to PIR Accession Number B41966"

/codon_start=1

/transl_table=11

/product="Virulence membrane protein phoQ."

/protein_id="BAA35951.1"

/db_xref="GI:1651557"

/translation="MKKLLRFFLSLVRFLATAAVVLVLSLAYGVALIGYSVSF DKTTPRLRGESNLFYTLAKWNNKLHVELPENIDKQSPMTLIYDENGQLLWAQRDV PMLKMIQDFDLKNSGFHEIADVNDTSLLSGDSHQQLQOEVEDDDDAEMTHSVA RVYEPATGRMKLTIVVDVTIPVELKSSYMWMSFYVLSANLLVLPILLVAAWMSL NNIEATGELEHREHRELLNPATRELTSLVRNLNLLKSERERYDKYRTTLDLT HSLKTPVAVLOSTLSLRSEKMSVDAEPVLEQISRTSQOIGYVLRASMRGGTLLS RELHPVADLONLTSALNKVQKGVNISLDSPEISFGEQNDVEYMGVNLDNACK YCLERFVLSAQTDDEHLIYVEDDGGPILSKREVFIDRGQRVDTLRPGQGVGLAVAR EITEQYEGKIVAGESMLGGMARMEVIFGRQHSAPKDE"

complement(6010..6681)

/gene="phoP"

complement(6010..6681)

/gene="phoP"

/note="ORF_ID:0240#1"

similar to SwissProt Accession Number P23836"

/codon_start=1

/transl_table=11

/product="transcriptional regulatory protein Phop."

/protein_id="BAA35952.1"

/db_xref="GI:1651558"

/translation="MRVLVEDNALLRHLHKVQIQDAGHQVDDAEDAEADYLLNEHI PDIAVLDGLDPEDEGLSLIRWRNSDVSPLIVLTARESMDKVEVLGAGADDYVTKP PHIERVMAQALMRNSGLASOVLSLPPQVDLSRRELSINDEVIKLTAFVTTMET LIRNGKVVSQDSLMLQLYPDALRESHTIDVLMGLRKKIKQAQIQPQEVITVRGQGY LFERL"

complement(6850..8220)

/gene="purB"

complement(6850..8220)

/gene="purB"

/note="ORF_ID:0240#2"

similar to PIR Accession Number S19212"

/codon_start=1

/transl_table=11

/product="Adenylosuccinate lyase (EC 4.3.2.2)"

/protein_id="BAA35953.1"

/db_xref="GI:1651559"

/translation="MELSSLTAVSPVDGRYGDVSKVYALRGIFSEYGLLFRVQVEVVRWL QKLAHAHAKEVPAPFAADIGYDAIVASFSEDAARIKTERTTNHVDVKAVEYFLKE KVAETPEIKAVSEFTHFACTSEDINNLHMLKLTARDEVLTPRQLTDGKDLAVQ YRDIPLLSRTHGQATPTIGKEMANVAYRMEROYRQLNOVEILGKINGAGVNAHI AAYPEVDNHQSEEFVTSIGIOWNPYTTQIEPHDYIAELFDCVARENTLILDFDRDW

/transl_table=11
/product="Hypothetical protein 23."
/protein_id="BAA35963.1"
/db_xref="GI:4062714"
/translation="MARNYDITLALAGICQARLVQOALHQHCHDADALHVSLSNII
DMNPSSTLAVFGSEANLVGLTLLGVNASSRQGLNAELTRYTLTSLMLERKLSSA
KALDTLGNRLNGRLEHFDLQSETLMSAMAAIYVDVIFSLPQIVTSGSPAVLOS
POVQAKVRATLACIRAAVLVHWHQVGGGRLQLMFNRRLTTQAKILAHLPPEL"
complement(3218..4324)
/gene="ycfB"
/complement(3218..4324)
/note="ycfB"
/gene="orf_ID:0240#4
similar to SwissProt Accession Number P25745"
/codon_start=1
/transl_table=11
/product="Hypothetical protein in purB 5' region (orf-15)
".
/protein_id="BAA35964.1"
/db_xref="GI:4062715"
/translation="MSETAKKVIIVGMSGVDSSVAMLLQOQGVQVEGLPMKNEEDD
GEECTAAADLAQAQVCDKGLIELHTVNFNAEYWDNVFELFAEYKAGTRPNDILC
NKEIKFALEFAEDLGDYATGHTVNRADVDGKRLRLGLSDNQDQSTFLTLSH
EQIAISLPFVGELEKPPQVKTAEDLGLVTAKKDSTGTCFIFGERKFFRGLYLPAAQ
KIIITVDGEIGEHQGLMYHTLQKRGKLGIGTGKTEGTEEPWYVVDKVDENNILVPAQ
HEHRLMSVGLIAQLHWDREPTGTMRCTVTRVQRTDIPCTVKALDDDDRIEVID
EPVAAVPGQSAVFYNGEVCGLGGIIEQRULPVV"
complement(4378..4839)
/note="orf_ID:0240#5
similar to PIR Accession Number S64074"
/codon_start=1
/transl_table=11
/product="Hypothetical protein YGL067"
".
/protein_id="BAA35965.1"
/db_xref="GI:4062716"
/translation="MFKPHVTVCVHAEKGLVVEETINGKALMNOPAGHLEADETL
VEAAALWEETGISAQOHPFIRHOMIAPDKTPFLFLFALELEQICPTOPHSDSID
CCRWSAEELIQASNLRSPLVAESIRCYQSQRYPLEMIGDNFPFKGVI"
complement(4849..5502)
/note="orf_ID:0240#6
similar to PIR Accession Number I64156"
/codon_start=1
/transl_table=11
/product="Hypothetical protein HI0694"
".
/protein_id="BAA35966.1"
/db_xref="GI:4062717"
/translation="MRQFIISENTMOKTSFRNHQKRFSSQRSTRRKPNOPTRVILF
NKPYDVLPPQFDEAGRTLKEFIPQGVYAAAGRLDRDSEGLLVLTNNALQARLTQPG
KRTGKIYVQVEGIPTDQALEALRNGVTLNDGPTLPAGAELVDEPAWLWPNRPPIRER
KSIPTSWIKIPLYEGRNQVRMTAHVGFPLRLIRYAMGDYSLDNLANGEWEVTD"
5674..6924
/gene="icd"
/note="icdA; icdE"
5674..6924
/gene="icd"
/note="orf_ID:0240#7
similar to SwissProt Accession Number P08200"
/codon_start=1
/transl_table=11
/product="Isocitrate dehydrogenase (NADP) (EC 1.1.1.42)
(oxalosuccinate decarboxylase) (icd) (NADP+specific icdh)
(icd)".
/protein_id="BAA35967.1"
/db_xref="GI:1651566"
/translation="MESKRVVPAQKKITLQKLNVPENPIPIYIEGDGIGVDVTPA
MLKVDAVERKAYGKISWMEIYTGKSTQVYQDWLPAETLDLIREYRAIKGP
LITPVGGRISLNALRQELDLYICLRVRYOQTSPVKHPDLTDMVIFRESEDIY
AGIEWKADSADAERKVIKFLREEMGVKIRFFEHCGIGIKPCSEGTGKRLVRAAEIAY

Query Match 3.4%; Score 23; DB 1; Length 20284;
Best Local Similarity 100.0%; Pred. No. 0.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 133 CTGTTCAATAAACCTACGATGT 155
|||||
Db 5376 CTGTTCAATAAACCTACGATGT 5354
|||||
RESULT 5
AC073744_0/C
WPCOMMENT
Sequence split into 5 fragments LOCUS AC073744 Accession AC073744
Fragment Name Begin End
AC073744_0 1 110000
AC073744_1 100001 210000
AC073744_2 200001 310000
AC073744_3 300001 410000
AC073744_4 400001 466859
LOCUS AC073744 466859 bp DNA linear HTG 12-JUL-2000
DEFINITION Mus musculus clone RP23-288A3, WORKING DRAFT SEQUENCE, 196
unordered pieces.
ACCESSION AC073744
VERSION AC073744.1 GI:8810361
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 466859)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 466859)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 1847467
Center clone name: RPCI-23_288A3

Summary Statistics
Consensus quality: 310234 bases at least Q40
Consensus quality: 387412 bases at least Q30
Consensus quality: 406374 bases at least Q20
Estimated insert size: 193000; pulse field gel estimation
Estimated insert size: 447359; sum-of-contigs estimation
Quality coverage: 6.42 in Q20 bases; pulse field gel estimation
Quality coverage: 2.77 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 196 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 1152: contig of 1152 bp in length
* 1153 1252: gap of unknown length
* 1253 2396: contig of 1144 bp in length
* 2397 2496: gap of unknown length
* 2497 4003: contig of 1507 bp in length
* 4004 4103: gap of unknown length
* 4104 5282: contig of 1179 bp in length
* 5283 5382: gap of unknown length
* 5383 6463: contig of 1081 bp in length
* 6464 6564: gap of unknown length
* 6564 7611: contig of 1048 bp in length
* 7611 7712: gap of unknown length
* 7712 9030: contig of 1319 bp in length
* 9031 9130: gap of unknown length

LOCUS	AP002555	222605 bp	DNA	linear	BCT 07-MAR-2001
DEFINITION	Escherichia coli O157:H7 DNA, complete genome, section 6/20.				
ACCESSION	AP002555	BA000007			
VERSION	AP002555.1	GI:13360886			
KEYWORDS					
SOURCE	Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952) DNA.				
ORGANISM	Escherichia coli O157:H7				
REFERENCE	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.				
AUTHORS	1 (sites)				
TITLE	Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S., Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T., Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and Shinagawa,H.				
	Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7				
	derived from the Sakai outbreak				
	Genes Genet. Syst. 74 (5), 227-239 (1999)				
	20198780				
JOURNAL MEDLINE	2 (sites)				
REFERENCE	Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M., Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and Hayashi,T.				
AUTHORS	Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an Escherichia coli K-12 strain MG1655				
TITLE	Syst. Appl. Microbiol. 23 (3), 315-324 (2000)				
	20557356				
	3 (sites)				
	Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S., Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T., Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and Shinagawa,H.				
	Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli O157:H7 strain derived from the Sakai outbreak				
JOURNAL MEDLINE	Gene 258 (1-2), 127-139 (2000)				
REFERENCE	20564182				
AUTHORS	4 (sites)				
TITLE	Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K., Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T., Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C., Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and Shinagawa,H.				
	Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12				
	DNA Res. 8 (1), 11-22 (2001)				
	5 (bases 1 to 222605)				
	Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and Hayashi,T.				
JOURNAL MEDLINE	Direct Submission				
REFERENCE	Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:ken@gen-info.osaka-u.ac.jp, URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365, Fax:81-6-6879-2047)				
AUTHORS	genome project.				
COMMENT	Location/Qualifiers				
	1. .222605				
	/organism="Escherichia coli O157:H7"				
	/strain="O157:H7"				
	/sub_strain="RIMD 0509952"				
FEATURES	/db_xref="taxon:83334"				
gene	74. .607				
	/gene="ECs1423"				
	74. .607				
	/gene="ECs1423"				
	/note="similar to B1045_ECOLI gi 1787283 percent identity 100 in 177 aa (Conserved in E.coli K-12)"				
CDS	/codon_start=1				
gene	/evidence=not_experimental				
	5086. .7659				
	/gene="ECs1427"				
	5086. .7659				
	/evidence=not_experimental				
gene	/transl_table=11				
	/product="putative polyprotein"				
	/protein_id="BAB34846.1"				
	/db_xref="GI:13360887"				
	/translation="MKTRIHHVVGQDITKLAVDVIVNAANPSLMGGGVGDAIHHRAAGP ALLDACLRVQQGDCPIGHAVITLAGDLPKAVVHTVGPVRRGEQNRDQLQDLYL NSLRVVAANSYTSVAFPAISTGVYGPRAAAAEIAVKTVEFITHALPEQYFVCYD EENAHLYERLLITQQQDE"				
CDS	549. .2030				
gene	/gene="ECs1424"				
	549. .2030				
	/gene="ECs1424"				
	/note="similar to YMDC_ECOLI gi 1787284 percent identity 99 in 493 aa (Conserved in E.coli K-12)"				
	/codon_start=1				
gene	/evidence=not_experimental				
	/transl_table=11				
	/product="putative synthase"				
	/protein_id="BAB34847.1"				
	/db_xref="GI:13360888"				
CDS	/translation="MMKKTPTSTKDSLPNKEMNDLPRLASAVLPLCSOHPGCGGLFPL EKSLDAFARYRLAEMSEHTLDVOYIYIQQDDMSGRLFSALLAAAKRGVYRLILLDDN NTPGLDDILLRDSHPRIEVLNPFSEFLRLPLGTYIDFSRLNRHMHKNSFTVDGVV TIVGGRTGDAYFGAGEEPLFSDLDVMAIGPVVEDVADDFARYWCYSVSPLOQVLDV PEGEMADRIELPASWGHNDAMTHRYLRKMESPFIINHLVDGTLPLIAKTRLLSDDPK GEGKARHSLPQRLFDIMGSPSERIDIISSYFVPTRAGVALLRMVRKGVKGTAILTN SLAANDVAVHAGYARWRKLLRYGVELYELKPTREOSTLHDRTGTGNSGASLHAKT FSIDKTYFIGSFNEDPSTLLNTMGFVIESETLAQLIDKRFIQSQYDAANQLRLDR WGRINWDRHAKKEIVLKKEPATSFWRKRVWRVRLASILPVEWLL"				
gene	complement(2038. .3195)				
CDS	/gene="ECs1425"				
gene	complement(2038. .3195)				
	/gene="ECs1425"				
	/note="similar to MDOC_ECOLI gi 3025226sp p75920 percent identity 99 in 385 aa (Conserved in E.coli K-12)"				
	/codon_start=1				
	/evidence=not_experimental				
gene	/transl_table=11				
	/product="glucans biosynthesis protein"				
	/protein_id="BAB34848.1"				
	/db_xref="GI:13360889"				
	/translation="MNPVPAQREYFLDSIRAWMLLGIPIHISLIYSSHTWVNSAEP SWLTLFNDFIHSFRMQVFFVSGYSYMLFRLPLKWKMKVVERVGIPLMTAIPLL TLPQFTIMQYVKGKASWPGSLYDKYNTLAWELISHLWFLVLVVMVTTLCVWIFKRI RNLNSDKTKKFSVMKLSVIFCLGIGYAVIRRTIFIVYPIILNSGTENFTVMOTL FYLPFILLGALAFIPHLKALFTPTPSRGCTLAALAFAYLLNORYSGSDAMWYTES VITWVGLMMVNVVFSFGHRLNFOSARVTFVNASLFIYLVHHPILTLFFGATITPHI TSNWLGFLGGLIFVVGIIAILYIHLRIPLLLKFLFSGKPVVKRENDKAPAR"				
gene	3588. .5123				
CDS	/gene="ECs1426"				
gene	3588. .5123				
	/note="similar to MDOG_ECOLI gi 1787286 percent identity 99 in 511 aa (Conserved in E.coli K-12)"				
	/codon_start=1				
	/evidence=not_experimental				
	/transl_table=11				
gene	/product="periplasmic glucans biosynthesis protein"				
	/protein_id="BAB34849.1"				
	/db_xref="GI:13360890"				
	/translation="MMKMRWLSAAVMLTLYTSSWAFSIDDVAKQAQSLAGKGYEAPK SNLPSVFRDMKYADYQIQFNHDKAYWNMLKTPFKLEFYHOGMYFDTVPKINEVTATA VKRIKSPDYTFGDVQHDKDTVDKDLGFAKVLVYIPNSKDNDEIVSMGLASVRYI GAGOVYGLSARGLAIDTALPSGEEFPPEKFEWIERKPTDKRLTIYALLDSPRATGAY KFYVMPGRDVTVDVQSKIYLRDKVKGLGVAPLTSMLFPGPNQSPANNRPRPELHDSNG LSIHAGNGEVIWRPLNPNKLAVSSFMENPQGLLQGRDPSFREDJDDRDYDRPS ALWTPKGEWKGKSELVEIPTNDETINDNIYAYTMDQLPEPKMKNFYITFSRDED KHTAPDNAMVQOTRRSTGDVKQSNLIRIQPDGTIAFVVDFTGAEMKPKDGTPTVTAQTS IGDNGEIVESTVRYNPVTKGRWRLVMRVKVKDAKKTETEMRAALVNADQTLTSETWSYQLP ANE"				
gene	5086. .7659				
CDS	/gene="ECs1427"				
gene	5086. .7659				
	/evidence=not_experimental				
	/transl_table=11				
	/product="putative polyprotein"				
	/protein_id="BAB34846.1"				
/db_xref="GI:13360887"					
/translation="MKTRIHHVVGQDITKLAVDVIVNAANPSLMGGGVGDAIHHRAAGP ALLDACLRVQQGDCPIGHAVITLAGDLPKAVVHTVGPVRRGEQNRDQLQDLYL NSLRVVAANSYTSVAFPAISTGVYGPRAAAAEIAVKTVEFITHALPEQYFVCYD EENAHLYERLLITQQQDE"					
CDS	549. .2030				
gene	/gene="ECs1424"				
	549. .2030				
	/gene="ECs1424"				
	/note="similar to YMDC_ECOLI gi 1787284 percent identity 99 in 493 aa (Conserved in E.coli K-12)"				
	/codon_start=1				
gene	/evidence=not_experimental				
	/transl_table=11				
	/product="putative synthase"				
	/protein_id="BAB34847.1"				
	/db_xref="GI:13360888"				
CDS	/translation="MMKKTPTSTKDSLPNKEMNDLPRLASAVLPLCSOHPGCGGLFPL EKSLDAFARYRLAEMSEHTLDVOYIYIQQDDMSGRLFSALLAAAKRGVYRLILLDDN NTPGLDDILLRDSHPRIEVLNPFSEFLRLPLGTYIDFSRLNRHMHKNSFTVDGVV TIVGGRTGDAYFGAGEEPLFSDLDVMAIGPVVEDVADDFARYWCYSVSPLOQVLDV PEGEMADRIELPASWGHNDAMTHRYLRKMESPFIINHLVDGTLPLIAKTRLLSDDPK GEGKARHSLPQRLFDIMGSPSERIDIISSYFVPTRAGVALLRMVRKGVKGTAILTN SLAANDVAVHAGYARWRKLLRYGVELYELKPTREOSTLHDRTGTGNSGASLHAKT FSIDKTYFIGSFNEDPSTLLNTMGFVIESETLAQLIDKRFIQSQYDAANQLRLDR WGRINWDRHAKKEIVLKKEPATSFWRKRVWRVRLASILPVEWLL"				
gene	complement(2038. .3195)				
CDS	/gene="ECs1425"				
gene	complement(2038. .3195)				
	/gene="ECs1425"				
	/note="similar to MDOC_ECOLI gi 3025226sp p75920 percent identity 99 in 385 aa (Conserved in E.coli K-12)"				
	/codon_start=1				
	/evidence=not_experimental				
gene	/transl_table=11				
	/product="glucans biosynthesis protein"				
	/protein_id="BAB34848.1"				
	/db_xref="GI:13360889"				
	/translation="MNPVPAQREYFLDSIRAWMLLGIPIHISLIYSSHTWVNSAEP SWLTLFNDFIHSFRMQVFFVSGYSYMLFRLPLKWKMKVVERVGIPLMTAIPLL TLPQFTIMQYVKGKASWPGSLYDKYNTLAWELISHLWFLVLVVMVTTLCVWIFKRI RNLNSDKTKKFSVMKLSVIFCLGIGYAVIRRTIFIVYPIILNSGTENFTVMOTL FYLPFILLGALAFIPHLKALFTPTPSRGCTLAALAFAYLLNORYSGSDAMWYTES VITWVGLMMVNVVFSFGHRLNFOSARVTFVNASLFIYLVHHPILTLFFGATITPHI TSNWLGFLGGLIFVVGIIAILYIHLRIPLLLKFLFSGKPVVKRENDKAPAR"				
gene	3588. .5123				
CDS	/gene="ECs1426"				
gene	3588. .5123				
	/note="similar to MDOG_ECOLI gi 1787286 percent identity 99 in 511 aa (Conserved in E.coli K-12)"				
	/codon_start=1				
	/evidence=not_experimental				
	/transl_table=11				
gene	/product="periplasmic glucans biosynthesis protein"				
	/protein_id="BAB34849.1"				
	/db_xref="GI:13360890"				
	/translation="MMKMRWLSAAVMLTLYTSSWAFSIDDVAKQAQSLAGKGYEAPK SNLPSVFRDMKYADYQIQFNHDKAYWNMLKTPFKLEFYHOGMYFDTVPKINEVTATA VKRIKSPDYTFGDVQHDKDTVDKDLGFAKVLVYIPNSKDNDEIVSMGLASVRYI GAGOVYGLSARGLAIDTALPSGEEFPPEKFEWIERKPTDKRLTIYALLDSPRATGAY KFYVMPGRDVTVDVQSKIYLRDKVKGLGVAPLTSMLFPGPNQSPANNRPRPELHDSNG LSIHAGNGEVIWRPLNPNKLAVSSFMENPQGLLQGRDPSFREDJDDRDYDRPS ALWTPKGEWKGKSELVEIPTNDETINDNIYAYTMDQLPEPKMKNFYITFSRDED KHTAPDNAMVQOTRRSTGDVKQSNLIRIQPDGTIAFVVDFTGAEMKPKDGTPTVTAQTS IGDNGEIVESTVRYNPVTKGRWRLVMRVKVKDAKKTETEMRAALVNADQTLTSETWSYQLP ANE"				
gene	5086. .7659				
CDS	/gene="ECs1427"				
gene	5086. .7659				
	/evidence=not_experimental				
	/transl_table=11				
	/product="putative polyprotein"				
	/protein_id="BAB34846.1"				
/db_xref="GI:13360887"					
/translation="MKTRIHHVVGQDITKLAVDVIVNAANPSLMGGGVGDAIHHRAAGP ALLDACLRVQQGDCPIGHAVITLAGDLPKAVVHTVGPVRRGEQNRDQLQDLYL NSLRVVAANSYTSVAFPAISTGVYGPRAAAAEIAVKTVEFITHALPEQYFVCYD EENAHLYERLLITQQQDE"					
CDS	549. .2030				
gene	/gene="ECs1424"				
	549. .2030				
	/gene="ECs1424"				
	/note="similar to YMDC_ECOLI gi 1787284 percent identity 99 in 493 aa (Conserved in E.coli K-12)"				
	/codon_start=1				
gene	/evidence=not_experimental				
	/transl_table=11				
	/product="putative synthase"				
	/protein_id="BAB34847.1"				
	/db_xref="GI:13360888"				
CDS	/translation="MMKKTPTSTKDSLPNKEMNDLPRLASAVLPLCSOHPGCGGLFPL EKSLDAFARYRLAEMSEHTLDVOYIYIQQDDMSGRLFSALLAAAKRGVYRLILLDDN NTPGLDDILLRDSHPRIEVLNPFSEFLRLPLGTYIDFSRLNRHMHKNSFTVDGVV TIVGGRTGDAYFGAGEEPLFSDLDVMAIGPVVEDVADDFARYWCYSVSPLOQVLDV PEGEMADRIELPASWGHNDAMTHRYLRKMESPFIINHLVDGTLPLIAKTRLLSDDPK GEGKARHSLPQRLFDIMGSPSERIDIISSYFVPTRAGVALLRMVRKGVKGTAILTN SLAANDVAVHAGYARWRKLLRYGVELYELKPTREOSTLHDRTGTGNSGASLHAKT FSIDKTYFIGSFNEDPSTLLNTMGFVIESETLAQLIDKRFIQSQYDAANQLRLDR WGRINWDRHAKKEIVLKKEPATSFWRKRVWRVRLASILPVEWLL"				
gene	complement(2038. .3195)				
CDS	/gene="ECs1425"				
gene	complement(2038. .3195)				
	/gene="ECs1425"				
	/note="similar to MDOC_ECOLI gi 3025226sp p75920 percent identity 99 in 385 aa (Conserved in E.coli K-12)"				
	/codon_start=1				
	/evidence=not_experimental				
gene	/transl_table=11				
	/product="glucans biosynthesis protein"				
	/protein_id="BAB34848.1"				
	/db_xref="GI:13360889"				
	/translation="MNPVPAQREYFLDSIRAWMLLGIPIHISLIYSSHTWVNSAEP SWLTLFNDFIHSFRMQVFFVSGYSYMLFRLPLKWKMKVVERVGIPLMTAIPLL TLPQFTIMQYVKGKASWPGSLYDKYNTLAWELISHLWFLVLVVMVTTLCVWIFKRI RNLNSDKTKKFSVMKLSVIFCLGIGYAVIRRTIFIVYPIILNSGTENFTVMOTL FYLPFILLGALAFIPHLKALFTPTPSRGCTLAALAFAYLLNORYSGSDAMWYTES VITWVGLMMVNVVFSFGHRLNFOSARVTFVNASLFIYLVHHPILTLFFGATITPHI TSNWLGFLGGLIFVVGIIAILYIHLRIPLLLKFLFSGKPVVKRENDKAPAR"				
gene	3588. .5123				
CDS	/gene="ECs1426"				
gene	3588. .5123				
	/note="similar to MDOG_ECOLI gi 1787286 percent identity 99 in 511 aa (Conserved in E.coli K-12)"				
	/codon_start=1				
	/evidence=not_experimental				
	/transl_table=11				
gene	/product="periplasmic glucans biosynthesis protein"				
	/protein_id="BAB34849.1"				
	/db_xref="GI:13360890"				
	/translation="MMKMRWLSAAVMLTLYTSSWAFSIDDVAKQAQSLAGKGYEAPK SNLPSVFRDMKYADYQIQFNHDKAYWNMLKTPFKLEFYHOGMYFDTVPKINEVTATA VKRIKSPDYTFGDVQHDKDTVDKDLGFAKVLVYIPNSKDNDEIVSMGLASVRYI GAGOVYGLSARGLAIDTALPSGEEFPPEKFEWIERKPTDKRLTIYALLDSPRATGAY KFYVMPGRDVTVDVQSKIYLRDKVKGLGVAPLTSMLFPGPNQSPANNRPRPELHDSNG LSIHAGNGEVIWRPLNPNKLAVSSFMENPQGLLQGRDPSFREDJDDRDYDRPS ALWTPKGEWKGKSELVEIPTNDETINDNIYAYTMDQLPEPKMKNFYITFSRDED KHTAPDNAMVQOTRRSTGDVKQSNLIRIQPDGTIAFVVDFTGAEMKPKDGTPTVTAQTS IGDNGEIVESTVRYNPVTKGRWRLVMRVKVKDAKKTETEMRAALVNADQTLTSETWSYQLP ANE"				
gene	5086. .7659				
CDS	/gene="ECs1427"				
gene	5086. .7659				
	/evidence=not_experimental				
	/transl_table=11				
	/product="putative polyprotein"				
	/protein_id="BAB34846.1"				
/db_xref="GI:13360887"					
/translation="MKTRIHHVVGQDITKLAVDVIVNAANPSLMGGGVGDAIHHRAAGP ALLDACLRVQQGDCPIGHAVITLAGDLPKAVVHTVGPVRRGEQNRDQLQDLYL NSLRVVAANSYTSVAFPAISTGVYGPRAAAAEIAVKTVEFITHALPEQYFVCYD EENAHLYERLLITQQQDE"					
CDS	549. .2030				
gene	/gene="ECs1424"				
	549. .2030				
	/gene="ECs1424"				
	/note="similar to YMDC_ECOLI gi 1787284 percent identity 99 in 493 aa (Conserved in E.coli K-12)"				
	/codon_start=1				
gene	/evidence=not_experimental				
	/transl_table=11				
	/product="putative synthase"				
	/protein_id="BAB34847.1"				
	/db_xref="GI:13360888"				
CDS	/translation="MMKKTPTSTKDSLPNKEMNDLPRLASAVLPLCSOHPGCGGLFPL EKSLDAFARYRLAEMSEHTLDVOYIYIQQDDMSGRLFSALLAAAKRGVYRLILLDDN NTPGLDDILLRDSHPRIEVLNPFSEFLRLPLGTYIDFSRLNRHMHKNSFTVDGVV TIVGGRTGDAYFGAGEEPLFSDLDVMAIGPVVEDVADDFARYWCYSVSPLOQVLDV PEGEMADRIELPASWGHNDAMTHRYLRKMESPFIINHLVDGTLPLIAKTRLLSDDPK GEGKARHSLPQRLFDIMGSPSERIDIISSYFVPTRAGVALLRMVRKGVKGTAILTN SLAANDVAVHAGYARWRKLLRYGVELYELKPTREOSTLHDRTGTGNSGASLHAKT FSIDKTYFIGSFNEDPSTLLNTMGFVIESETLAQLIDKRFIQSQYDAANQLRLDR WGRINWDRHAKKEIVLKKEPATSFWRKRVWRVRLASILPVEWLL"				
gene	complement(2038. .3195)				
CDS	/gene="ECs1425"				
gene	complement(2038. .3195)				
	/gene="ECs1425"				
	/note="similar to MDOC_ECOLI gi 3025226sp p75920 percent identity 99 in 385 aa (Conserved in E.coli K-12)"				
	/codon_start=1				
	/evidence=not_experimental				
gene	/transl_table=11				
	/product="glucans biosynthesis protein"				
	/protein_id="BAB34848.1"				
	/db_xref="GI:13360889"				
	/translation="MNPVPAQREYFLDSIRAWMLLGIPIHISLIYSSHTWVNSAEP SWLTLFNDFIHSFRMQVFFVSGYSYMLFRLPLKWKMKVVERVGIPLMTAIPLL TLPQFTIMQYVKGKASWPGSLYDKYNTLAWELISHLWFLVLVVMVTTLCVWIFKRI RNLNSDKTKKFSVMKLSVIFCLGIGYAVIRRTIFIVYPIILNSGTENFTVMOTL FYLPFILLGALAFIPHLKALFTPTPSRGCTLAALAFAYLLNORYSGSDAMWYTES VITWVGLMMVNVVFSFGHRLNFOSARVTFVNASLFIYLVHHPILTLFFGATITPHI TSNWLGFLGGLIFVVGIIAILYIHLRIPLLLKFLFSGKPVVKRENDKAPAR"				
gene	3588. .5123				
CDS	/gene="ECs1426"				
gene	3588. .5123				
	/note="similar to MDOG_ECOLI gi 1787286 percent identity 99 in 511 aa (Conserved in E.coli K-12)"				
	/codon_start=1				
	/evidence=not_experimental				
	/transl_table=11				
gene	/product="periplasmic glucans biosynthesis protein"				
	/protein_id="BAB34849.1"				
	/db_xref="GI:13360890"				
	/translation="MMKMRWLSAAVMLTLYTSSWAFSIDDVAKQAQSLAGKGYEAPK SNLPSVFRDMKYADYQIQFNHDKAYWNMLKTPFKLEFYHOGMYFDTVPKINEVTATA VKRIKSPDYTFGDVQHDKDTVDKDLGFAKVLVYIPNSKDNDEIVSMGLASVRYI GAGOVYGLSARGLAIDTALPSGEEFPPEKFEWIERKPTDKRLTIYALLDSPRATGAY KFYVMPGRDVTVDVQSKIYLRDKVKGLGVAPLTSMLFPGPNQSPANNRPRPELHDSNG LSIHAGNGEVIWRPLNPNKLAVSSFMENPQGLLQGRDPSFREDJDDRDYDRPS ALWTPKGEWKGKSELVEIPTNDETINDNIYAYTMDQLPEPKMKNFYITFSRDED KHTAPDNAMVQOTRRSTGDVKQSNLIRIQPDGTIAFVVDFTGAEMKPKDGTPTVTAQTS IGDNGEIVESTVRYNPVTKGRWRLVMRVKVKDAKKTETEMRAALVNADQTLTSETWSYQLP ANE"				
gene	5086. .7659				
CDS	/gene="ECs1427"				
gene	5086. .7659				
	/evidence=not_experimental				
	/transl_table=11				
	/product="putative polyprotein"				
	/protein_id="BAB34846.1"				
/db_xref="GI:13360887"					
/translation="MKTRIHHVVGQDITKLAVDVIVNAANPSLMGGGVGDAIHHRAAGP ALLDACLRVQQGDCPIGHAVITLAGDLPKAVVHTVGPVRRGEQNRDQLQDLYL NSLRVVAANSYTSVAFPAISTGVYGPRAAAAEIAVKTVEFITHALPEQYFVCYD EENAHLYERLLITQQQDE"					
CDS	549. .2030				
gene	/gene="ECs1424"				
	549. .2030				
	/gene="ECs1424"				
	/note="similar to YMDC_ECOLI gi 1787284 percent identity 99 in 493 aa (Conserved in E.coli K-12)"				
	/codon_start=1				
gene	/evidence=not_experimental				
	/transl_table=11				
	/product="putative synthase"				
	/protein_id="BAB34847.1"				
	/db_xref="GI:13360888"				
CDS	/translation="MMKKTPTSTKDSLPNKEMNDLPRLASAVLPLCSOHPGCGGLFPL EKSLDAFARYRLAEMSEHTLDVOYIYIQQDDMSGRLFSALLAAAKRGVYRLILLDDN NTPGLDDILLRDSHPRIEVLNPFSEFLRLPLGTYIDFSRLNRHMHKNSFTVDGVV TIVGGRTGDAYFGAGEEPLFSDLDVMAIGPVVEDVADDFARYWCYSVSPLOQVLDV PEGEMADRIELPASWGHNDAMTHRYLRKMESPFIINHLVDGTLPLIAKTRLLSDDPK GEGKARHSLPQRLFDIMGSPSERIDIISSYFVPTRAGVALLRMVRKGVKGTAILTN SLAANDVAVHAGYARWRKLLRYGVELYELKPTREOSTLHDRTGTGNSGASLHAKT FSIDKTYFIGSFNEDPSTLLNTMGFVIESETLAQLIDKRFIQSQYDAANQLRLDR WGRINWDRHAKKEIVLKKEPATSFWRKRVWRVRLASILPVEWLL"				
gene	complement(2038. .3195)				
CDS	/gene="ECs1425"				
gene	complement(2038. .3195)				
	/gene="ECs1425"				
	/note="similar to MDOC_ECOLI gi 3025226sp p75920 percent identity 99 in 385 aa (Conserved in E.coli K-12)"				
	/codon_start=1				
	/evidence=not_experimental				
gene	/transl_table=11				
	/product="glucans biosynthesis protein"				
	/protein_id="BAB34848.1"				
	/db_xref="GI:13360889"				
	/translation="MNPVPAQREYFLDSIRAWMLLGIPIHISLIYSSHTWVNSAEP SWLTLFNDFIHSFRMQVFFVSGYSYMLFRLPLKWKMKVVERVGIPLMTAIPLL TLPQFTIMQYVKGKASWPGSLYDKYNTLAWELISHLWFLVLVVMVTTLCVWIFKRI RNLNSDKTKKFSVMKLSVIFCLGIGYAVIRRTIFIVYPIILNSGTENFTVMOTL FYLPFILLGALAFIPHLKALFTPTPSRGCTLAALAFAYLLNORYSGSDAMWYTES VITWVGLMMVNVVFSFGHRLNFOSARVTFVNASLFIYLVHHPILTLFFGATITPHI TSNWLGFLGGLIFVVGIIAILYIHLRIPLLLKFLFSGKPVVKRENDKAPAR"				
gene	3588. .5123				
CDS	/gene="ECs1426"				
gene	3588. .5123				
	/note="similar to MDOG_ECOLI gi 1787286 percent identity 99 in 511 aa (Conserved in E.coli K-12)"				
	/codon_start=1				
	/evidence=not_experimental				
	/transl_table=11				
gene	/product="periplasmic glucans biosynthesis protein"				
	/protein_id="BAB34849.1"				
	/db_xref="GI:13360890"				
	/translation="MMKMRWLSAAVMLTLYTSSWAFSIDDVAKQAQSLAGKGYEAPK SNLPSVFRDMKYADYQIQFNHDKAYWNMLKTPFKLEFYHOGMYFDTVPKINEVTATA VKRIKSPDYTFGDVQHDKDTVDKDLGFAKVLVYIPNSKDNDEIVSMGLASVRYI GAGOVYGLSARGLAIDTALPSGEEFPPEKFEWIERKPTDKRLTIYALLDSPRATGAY KFYVMPGRDVTVDVQSKIYLRDKVKGLGVAPLTSMLFPGPNQSPANNRPRPELHDSNG LSIHAGNGEVIWRPLNPNKLAVSSFMENPQGLLQGRDPSFREDJDDRDYDRPS ALWTPKGEWKGKSELVEIPTNDETINDNIYAYTMDQLPEPKMKNFYITFSRDED KHTAPDNAMVQOTRRSTGDVKQSNLIRIQPDGTIAFVVDFTGAEMKPKDGTPTVTAQTS IGDNGEIVESTVRYNPVTKGRWRLVMRVKVKDAKKTETEMRAALVNADQTLTSETWSYQLP ANE"				
gene	5086. .7659				
CDS	/gene="ECs1427"				
gene	5086. .7659				
	/evidence=not_experimental				
	/transl_table=11				
	/product="putative polyprotein"				
	/protein_id="BAB34846.1"				
/db_xref="GI:13360887"					
/translation="MKTRIHHVVGQDITKLAVDVIVNAANPSLMGGGVGDAIHHRAAGP ALLDACLRVQQGDCPIGHAVITLAGDLPKAVVHTVGPVRRGEQNRDQLQDLYL NSLRVVAANSYTSVAFPAISTGVYGPRAAAAEIAVKTVEFITHALPEQYFVCYD EENAHLYERLLITQQQDE"					
CDS	549. .2030				
gene	/gene="ECs1424"				
	549. .2030				
	/gene="ECs1424"				
	/note="similar to YMDC_ECOLI gi 1787284 percent identity 99 in 493 aa (Conserved in E.coli K-12)"				
	/codon_start=1				

/gene="ECs1427"
/note="similar to membrane glycosyltransferase MDOH_ECOLI
g1|1787287 percent identity 100 in 847 aa but differs at
N-ter (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="membrane glycosyltransferase"
/protein_id="BAB34850.1"
/db_xref="GI:13360891"
/translation="MKPGATSYLPNNKTTEYIDAMPIAASEKAALPKTDIRAVHQALD
AERTWAREDDSPQGSVKARLEQANPDSIADQLIKDDRGDLKAMPKRSMPFD
PRTNPGVRENRGRDQVTPRLAKLTKEEQESOKWRTVGTIRYRIILLITLAQTV
VATWMTKTLIPQGWALINPMDVQGLPMLPYMLQGTGILIFAVLFCWSAG
FTWMLGTLILGHRDKYSISATVSGDEPLNPEHRTALIMPICNEDVNRVFAGLRWAT
ESVKATGNKHEFDVILSDSYNPDCVABQKAWMELIAEVGEGGQIFYYRRRRRVKTK
SGNIDFCRRWGSQSYMYVLDADSVMTGDCICGLVLRMEANPNAGIIOSSPKASGMD
TLYARCOQFATRVYGPLEFAGLHFQWLGESHYWGNAIIRVKPFFIEHCALAPLPGEGS
FAGSILSHDFVEAALMRAGWGVWIAIDLPGSYEELPPNLDDELKDRRWCHGNLMNF
RLFLVGMHPVHRVFLTGVMYSILSAPUMFALSTALQVVALTEPOYFLQPRQLF
PVMQWRPELAIALFASMTVLLFLPKLLSILLIWKCKTKEYGGFWRTLSLLLEVLFS
VLLAPVRLMFLHFVYVSAFLGWGVVWNSPQRDDSTSWGEAFKRHGSOLLISGLVWAVG
MAWLDLRFELWLAFTVFSLISPFVSVISSRATVGLRTKRKLELIPREYSPQVILD
TQRELMNRQSLDGMHAFVNPENALATAMATARHASKVLEIARDRHVEQALNE
TPEKLNRRRLVLLSDPYTMARLHFRVWNSPERISWSVSIYEGIKLNPLALRKPDAAS
Q"
7832. .8059
/gene="ECs1428"
7832. .8059
/gene="ECs1428"
/note="identical to YCEK_ECOLI g1|1787288 (Conserved in
E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB34851.1"
/db_xref="GI:13360892"
/translation="MRLIVSIVMTVLLSGCGSIISRTIPGQGHGNYYPGVQWQVDRDS
AMRYVTILDLPFSLVFDLTLLPIDIHGPGYE"
complement(8060. .8437)
/gene="ECs1429"
complement(8060. .8437)
/gene="ECs1429"
/note="similar to MSYB_ECOLI g1|1787289 percent identity
99 in 125 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="acidic protein MsyB"
/protein_id="BAB34852.1"
/db_xref="GI:13360893"
/translation="MMTMVATLEEAIDAAREEFILADNPGLDAEDANVQOOFNAQYVILQ
DGDIMVQVEFFADEGECEGLPMLSGEAAQSVFDGDDYDEIRQEWQENTLHWEDEG

Query Match 3.4%; Score 23; DB 1; Length 222605;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 CTGTTCAATAAACCCCTACGATGT 155
|||||
Db 151662 CTGTTCAATAAACCCCTACGATGT 151640

RESULT 7
AE008754/c
LOCUS Salmonella typhimurium LT2, section 58 of 220 of the complete
DEFINITION genome.
ACCESSION AE008754 AE006468
VERSION AE008754.1 GI:16419750
KEYWORDS
SOURCE Salmonella typhimurium LT2.

Salmonella typhimurium LT2
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmonella
1 (bases 1 to 24578)
McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
Stonking,T., Nhan,M., Waterston,R. and Wilson,R.K.
Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2
Nature 413 (6858), 852-856 (2001)
21534948
11677609
2 (bases 1 to 24578)
The Salmonella typhimurium Genome Sequencing Project.
Direct Submission
Submitted (29-MAR-2001) Genome Sequencing Center, Department of
Genetics, Washington University School of Medicine, 4444 Forest
Park Boulevard, St. Louis, MO 63108, USA
COMMENT Supported by NIH grant 5U 01 AI43283

Coding sequences below are predicted from manually evaluated
computer analysis, using similarity information and the programs;
GLIMMER; <http://www.tigr.org/softlab/glimmer/glimmer.html> and
GeneMark; <http://opal.biology.gatech.edu/GeneMark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>,
and Pedro Romero and Peter Karp at EcoCyc;
<http://ecocyc.PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites
were kindly provided by Heladia Salgado, Julio Collado-Vides and
ReguonDB;
http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistries or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one ml3 subclone.

Location/Qualifiers
1. 24578
/organism="Salmonella typhimurium LT2"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
/note="LT2"
complement(87. .1472)
/gene="purB"
/note="STMI232"
complement(87. .1457)
/gene="purB"
/EC_number="4.3.2.2"
/note="similar to E. coli adenylosuccinate lyase
(AAC74215.1); Blastp hit to AAC74215.1 (456 aa), 94%
identity in aa 1 - 456"
/codon_start=1
/transl_table=11
/product="adenylosuccinate lyase"
/protein_id="AAL20161.1"
/db_xref="GI:16419751"
/translation="MELSSLTAVSPVDGRYGDVKVSALRGIFSEYGLLKFRVQVEVRWL
QKLAHAAIKVEPAFAADANGYLDLVANFEADAARIKTIERTNHDVRAVEYFLKE
KVAAIPALHDSVFSEIHPACTSEDINNLSHALMLKLTARDEVILPWROVINAVKDLATQ
YRDIPLLRTHGPGATPSTLGKMANVAYRERQFROLNQVEILKINGAVGNNAHI
AAYPEVDHGFSEFVTSLGIQWNPYTTQIEPDHYIAELFDCIARFNILIDPDQVW
GYIALNHFKOKTAGESITMPHKVNPIDFENSEGNLSNVLHHLANKLKPVSRWQ
RLDUTSTVLRNLGVIGYALTATYQSLTKGVSKLEVNDRDLDELHDHNEVLAEP IQTV
MRRYGIETKPYEKLKELTRGKRVDAGMKQFIDSLALPEAEKTRLKAMTPANYIGRAVT
LVDELK"

gene	complement(1461. .2160) /gene="ycfc" /note="STM1233" complement(1461. .2108) /gene="ycfc" /note="similar to E. coli orf, hypothetical protein (AAC74216.1); Blastp hit to AAC74216.1 (213 aa), 94% identity in aa 1 - 213" /codon_start=1 /transl_table=11 /product="unknown membrane associated protein" /protein_id="AAL20162.1" /db_xref="GI:16419752" /translation="MIVAKNYDITLALSGICQARLVQQLAHOGHCDADALHVSLSN VIDNPSSTGLVFGSGSEANRLGLETLILGVLNASSRQGLNAELTRYLSLMVLSKLS SAKGALNTLGRINGLQRLDHFLEQSDTLMSAMAGIYVDVSPRIQVITGSPAVL QSPQVQARVSLAGIRAAVLMHQVGGRLQLMFSRHLRTTQAKILAHLTPEL" complement(1467. .1472) /gene="ycfc" /note="putative RBS for purB; RegulonDB:STMS1H001488" complement(2112. .2117) /gene="ycfc" /note="putative RBS for ycfC; RegulonDB:STMS1H001489" complement(2134. .2142) /gene="ycfc" /note="putative -10 signal for ycfC; RegulonDB:STMLTH004479" complement(2152. .2160) /gene="ycfc" /note="putative -35 signal for ycfC; RegulonDB:STMLTH004479" complement(2189. .3340) /gene="trmU" /note="STM1234" complement(2189. .3340) /gene="trmU" /EC_number="2.1.1.61" /note="similar to E. coli orf, hypothetical protein (AAC74217.1); Blastp hit to AAC74217.1 (383 aa), 94% identity in aa 1 - 383" /codon_start=1 /transl_table=11 /product="tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase" /protein_id="AAL20163.1" /db_xref="GI:16419753" /translation="MIEYALKFNWASIPMSPEPKVIVGMSGVDSSVSAWLQQQG YQVGLFMKNEEDGEYCTAAADLADAQVCDKLGIELHTVNFAAEYDNDVFEFL EYKAGTTPNPDIILCKEIKFKALEFAEDLGADYIATGHYVRRADVNGKSRLRLGL DGNKQDSFYLYTLGHEQIAQSLFPVGELEKPOVRKIAEDLGLVTAKKDGTCIGICE RKRFDFGRYLPAPQGRKITVDGDEIGEHOGLMYHTLGQRKGLGIGCTKDGSEDPWV VDKDENNVLLVAGGHEPRLMSVGLIAQLHWVDREFFTGTLCTYKTRYRQTDIFC TINLANDREIVFDEPVAAPVPGQSAVFYSGEVLGGGIIIEQLPLTV" complement(3344. .3349) /gene="trmU" /note="putative RBS for trmU; RegulonDB:STMS1H001490" complement(3349. .3820) /gene="ymfB" /note="STM1235" complement(3349. .3810) /gene="ymfB" /note="similar to E. coli putative phosphohydrolase (AAC74218.1); Blastp hit to AAC74218.1 (153 aa), 84% identity in aa 1 - 152" /codon_start=1 /transl_table=11 /product="putative MutT-like protein" /protein_id="AAL20164.1" /db_xref="GI:16419754" /translation="MFKPHVTVACIVHAEDKFLVVEETINGKSLWNOPAGHLEADETL AQAREALWETGTGAQPOFHWIAPDKTFFLRFALFKLDHTCAIDPHDNDID CCRWVSADFLNPLRSLPLVAESVRSYQTQGRYPILSVGEFNWPFTEGVK" complement(3815. .3820)
RBS	
gene	/gene="ymfB" /note="Putative RBS for ymfB; RegulonDB:STMS1H001491" complement(3822. .4159) /gene="STM1236" complement(3822. .4151) /gene="STM1236" /codon_start=1 /transl_table=11 /product="putative periplasmic protein" /protein_id="AAL20165.1" /db_xref="GI:16419755" /translation="MKRLITAFALLITPLSVNAASFDCQAKQAADKAICAHLTLDKDK VEMTKYQFLKGLFAMSGRGALDAQQSWLKQROCKRADVTCLTKAYNERLKQLDVIY DHIDKPL" complement(4148. .4813) /gene="ymfC" /note="STM1237" complement(4148. .4813) /gene="ymfC" /note="similar to E. coli orf, hypothetical protein (AAC74219.1); Blastp hit to AAC74219.1 (207 aa), 83% identity in aa 1 - 205" /codon_start=1 /transl_table=11 /product="putative ribosomal large subunit pseudouridine synthase" /protein_id="AAL20166.1" /db_xref="GI:16419756" /translation="MRQLISSENTMOKTSFRNHYVKRFSRQASKSRKENQPKRVVLF NKPYDVLPOFTDEAGRTLKDFIPQGVYAAGRLDRDSEGLLVLTNDGALQARLTQPG KRTGKIYVQVEGIPDAAALQALRTGVLNDGPTLPAGIEIVAEPPDWLPPTPIRER KNIPTSWLKVLYEGERNRQVRMTAHVGHPTLRLIRYSMGDTYTLNGDNGQWRIAE KDR" complement(4154. .4159) /gene="STM1236" /note="putative RBS for STM1236; RegulonDB:STMS1H001492" 4740. .4754 /note="putative binding site for FruR, RegulonDB: STMS1H00184" /bound_moiety="FruR" 4829. .6235 /gene="icdA" /note="STM1238" 4829. .4837 /gene="icdA" /note="putative -35 signal for icdA; RegulonDB:STMLTH004480"
RBS	
protein_bind	Query Match 3.3%; Score 22; DB 1; Length 24578; Best Local Similarity 100.0%; Pred. No. 0.8; Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
gene	Qy 71 ACCGATTCAGTTCACGACAAGC 92 Db 4749 ACCGATTCAGTTCACGACAAGC 4728
-35_signal	RESULT 8 AL627269/c LOCUS DEFINITION AL627269 Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18, complete chromosome; segment 5/20. ACCESSION AL627269 AL513382 VERSION AL627269.1 GI:16502231 SOURCE ORGANISM Salmonella enterica subsp. enterica serovar Typhi. Salmonella enterica subsp. enterica serovar Typhi Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella. REFERENCE 1 (bases 1 to 254050) AUTHORS Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D., Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G., Sebahia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,

Connerton, P., Cronin, A., Davis, P., Davies, R. M., Dowd, L., White, N., Farrar, J., Feltwell, T., Hamlin, N., Haque, A., Hien, T. T., Holroyd, S., Jagels, K., Krogh, A., Larsen, T. S., Leather, S., Moule, S., O'Gaora, P., Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S. and Barrell, B. G.
Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi Crl8
Nature 413 (6858), 848-852 (2001)
11677608
2 (bases 1 to 254050)
Parkhill, J.
Direct Submission
Submitted (25-OCT-2001) Submitted on behalf of the *Salmonella* sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK
E-mail: parkhill@sanger.ac.uk

COMMENT

Notes:
Details of *S. typhi* sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S_typhi/).

FEATURES

source
Location/Qualifiers
1. .254050
/organism="Salmonella enterica subsp. enterica serovar Typhi"
/strain="Crl8"
/db_xref="taxon:90370"
181. .2793
/gene="STY1078"
/note="pepN"
181. .2793
/gene="STY1078"
/note="Orthologue of *E. coli* pepN (AMPN_ECOLI); Pasta hit to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
/codon_start=1
/transl_table=11
/product="aminopeptidase N"
/protein_id="CAD08183.1"
/db_xref="GI:16502232"
/db_xref="SPTREMBL:Q827T0"
/translation="MTQQPQAKYRHDYRAPDYQITDIELTFDLDAEKTVVTAISQAVR HGADPLRLDGEDLTLVSHVNDAPMTAYKEEGALIIISDLPERFTLRVNIISPAARVAQEGLYSGDALCTQCEAGFRHITWYLDPRDVLARETTKIIADKSKYFELISNGN NLAQLELGRHWQWQDPPKPCYLFALVAGDFDLVRLDTFTTFRSGREVALEYVDGRG NLRDAPWMTSLANSKWDETREGLEYDLDIYMIYVADFPNMGAMENKGLNIFNSXVY LARPTATPKYLDLIERVIGHEYEHNTGNNRTCTRDWFQLSLEKGLTVRDQFESSDL GSRAVRNINVRTMRGLQFAEDASPMHPIRPDKVTEMNNFTLTIVKGAEVIRMIH TLLEENFQGMQYIFERHDGSAATCDDFVQAMEDASNDLSHFRWYQSQGPVIVY KDDYNPETEQYTTIISQRTPATDAQEKQPLHIPFAIELYDNEGNIPIQKGGHPVNA VLVNTQAEQTFITFDNVYFQVPALLCFEAPLVKLEYKWSQDLTFLMRHARNDFSRWD AQOSLLATYIKLVARHQOGQPLSLPVHVADAFRAVLLOEKIDPALAAETLTLPISANE IAELEVIDPIATAQVREALTRTIAELADEFIAYINANHLDYRYVDHGDIGKRTLRN ACURKLTFTGELANTLVSKQYRDANNMTDALAASAAVAQULPCRDITLMQYEDKRW H QDLGVMDKFMFLQSTSPAEVNLFTVRSLLKHSFMSNPNRISLIGAFAGSNPAAFH AQDGSQYQFVLEMTLNSRNPQASRLIEPLRLKRYDDKKRQEMRAALEQLKGLLEN LSGDLQKFTIKALA"
532. 1332
/gene="STY1078"
/note="Pfam match to entry PF01433 Peptidase_M1, Peptidase family M1, score 245.10, E-value 1.9e-84"
1060. .1089
/gene="STY1078"
/note="PS00142 Neutral zinc metallopeptidases, zinc-binding region signature"
3001. .4011
/gene="STY1079"
/note="pyrD"
3001. .4011
/gene="STY1079"
/note="Orthologue of *E. coli* pyrD (PYRD_ECOLI); Pasta hit to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
/codon_start=1
/transl_table=11

/product="dihydroorotate dehydrogenase"
/protein_id="CAD08184.1"
/db_xref="GI:16502233"
/db_xref="SPTREMBL:Q827S9"
/translation="MYYPVRKALFOLDPERAHEFTFOQLRRITGTPLEALVRQKYPT KPYTCMGLTFFKNPLGIAAGLDKDCIDALGAMGFGSLEIGTIVTPRQSGNDKRPURFR LVDAEGLINRMGNNLVDNLVENVKAKHFDGLGINIKNKDTPVENGKDDYLLICME KYVAYAGYIAINISSPTPLRTLOYGDALDDLLTAIKNKNDLQVYIHHKYVPVAVKI APDLCEELIQVADSLLRHNDIGVIATNTTLDRLSLVQGMKNCQOTGGLSGRPLQAKST EIIRLSLELKGQLPIIGVGDSIVIAAREKIANAGATLVQIYSGFIFKGPPLKEIVT HI"
3136. .3990
/gene="STY1079"
/note="Pfam match to entry PF01180 DHODEHASE, Dihydroorotate dehydrogenase, score 563.50, E-value 1.4e-165"
3238. .3297
/gene="STY1079"
/note="PS00911 Dihydroorotate dehydrogenase signature 1"
3874. .3936
/gene="STY1079"
/note="PS00912 Dihydroorotate dehydrogenase signature 2"
4174. .4719
/gene="STY1080"
4174. .4719
/gene="STY1080"
/note="STY1080"
/note="Orthologue of *E. coli* YCBW_ECOLI; Fasta hit to YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note discrepancy in position of translational start site."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08185.1"
/db_xref="GI:16502234"
/db_xref="SPTREMBL:Q827S8"
/translation="MMRIKPPDDNRWYIDEEDHRMLDLANGMLFRSFRSKMLTPDA FCPITGFCVDDAALYFSEBKCRDFELTKQRAELVLNALVAIRYLKQPMKSWHFVAH GEMWTPGTDAASVWLSDTAEQVNLVVEPGENAALLCLLAQPGVIAGRTMQLGDAIK INMDRLKPOVHCSESLEQAV"
complement(4716. .5825)
/gene="STY1081"
complement(4716. .5825)
/gene="STY1081"
/note="Orthologue of *E. coli* P75863; Fasta hit to P75863 (369 aa), 86% identity in 367 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08186.1"
/db_xref="GI:16502235"
/db_xref="SPTREMBL:Q827S7"
/translation="MATLSRLFHPVKSMRGIGLTHALADISGLAFDRIFMWTESDGT FITARQFQVRVFTSPSLHDLHLTAPDGSALVRFRDFTPDQAPTEVGNHETARVA PTAINQMLGSEFTSRDVLQWGVGQPLTRVRKHNALVFRFADGYPYLLTNEASLRDQQ RCPAGVQMQRFLNVLVSGVAAVEEDSKVLRIGVDFDVVKPCSCICFTTVPSPERQO KHPSGEPLATLOAFRTAODNGDVDFGONLIARNSGVIRVGVDEIILATAPAKAYGATT LODSVTPKHPDGSVTIDMQGQTFCGNNQOVLLEOLENQIRIPYSCRAGICCCCRIR LLEGEVSPLKKSAMGDDGTILSCSCVPKTAALLEN"
complement(4737. .4955)
/gene="STY1081"
/note="Pfam match to entry PF00111 fer2, 2Fe-2s iron-sulfur cluster binding domains, score 55.90, E-value 8.5e-13"
5924. .8032
/gene="STY1082"
5924. .8032
/gene="STY1082"
/note="Similar to Haemophilus influenzae hypothetical protein h10116 H10116 SW:YCBY_HAEIN (P44524; P43945) fasta scores: E(): 0. 60.3% id in 710 aa
Orthologue of *E. coli* YCBY_ECOLI; Fasta hit to YCBY_ECOLI (702 aa), 92% identity in 702 aa overlap"
/codon_start=1

misc_feature

gene

CDS

/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08187.1"
/db_xref="GI:16502236"
/db_xref="SPTREMBL:O82756"
/translation="MNSLFASARGLEELIKLEKLGAVCGVQGGVFHQFQDTRLI
YSLWGRSLAKRIIPGEGCKYSDDLIFGVQAINWTEIFNPGATFAVHFGSLNDLI
RNSQYGAMKVDAIVDAFTRNLPNPVRESPLRLINWLNKETIASIDSGDLGH
LRGYDRDTGLAPIKETLAIAVMRSWGQPGTLLDPMGCSGTSLIEAAMWATDRAPGL
HRHGWFSGAHQDEAIWEVKADAOFRARKGLAEYSSEHYGSGLSDARVTERASNAK
RAGIGELITRPVKDVAOLSNPKPYGTVISNPPYGERLDSPALIALHSLGLRTMK
NQFGWMLISASPDLGLSLQLRADQKFKNKGPLDCQVKNYHIAETLDSKPPIYA
EDIANRLKLNKLKWARQEGEICYLDADLPETYNADVRYGDWAWIQEYAPPKT
DAQKARLEFDITAAITSLVLGIPNNKLVLTREQRKNGQYQMSKEGGEFLVESEYNA
RLWNLTDTGLDIHTARRMLKMSGKDNLNFYSYTGSSASVHAGILGARNTTT
VMSRSTYLWEARNLGLSGRAHLRIOADCLGWLRANEQDFLIDPPTFSNSKR
MEESQVDORHVLMKDLKRLRRGGTIMFSNNKRGFRMDLEGELGLTAQEITOKT
LSPDFARNQHHCWLRRA"

misc_feature
/gene="STY1082"
/note="Pfam match to entry PF01170 UPF0020,
Uncharacterized protein family UPF0020, score 781.30,
E-value 3.7e-231"

misc_feature
/gene="STY1082"
/note="PS01261 Uncharacterized protein family UPF0020
signature"
/gene="STY1083"
/gene="STY1083"
/gene="STY1083"
/note="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in
526 aa overlap
Fasta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa
overlap
Orthologue of E. coli UUP_ECOLI; Fasta hit to UUP_ECOLI
(635 aa), 94% identity in 634 aa overlap"
/transl_table=11
/product="ABC transporter ATP-binding protein"

Query Match 3.3%; Score 22; DB 1; Length 254050;
Best Local Similarity 100.0%; Pred.No. 1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 AGCGATTACGCTCAGCAAGC 92
|||||

Db 180634 AGCGATTACGCTCAGCAAGC 180613

RESULT 9
AE013675
LOCUS AE013675 10393 bp DNA linear BCT 26-JUL-2002
DEFINITION Versinia pestis KIM section 75 of 415 of the complete genome.
ACCESSION AE013675 AE009952
VERSION AE013675.1 GI:21957454
KEYWORDS
SOURCE Versinia pestis KIM.
ORGANISM Versinia pestis KIM
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Versinia.

REFERENCE
1 (bases 1 to 10393)
AUTHORS Deng,W., Burland,V., Plunkett,G. III, Boutin,A., Mayhew,G.F.,
Liss,P., Perna,N.T., Rose,D.J., Mau,B., Zhou,S., Schwartz,D.C.,
Fetherston,J.D., Lindler,L.E., Brubaker,R.R., Plana,G.V.,
Straley,S.C., McDonough,K.A., Nilles,M.L., Matson,J.S.,
Blattner,F.R. and Perry,R.D.
Genome Sequence of Versinia pestis KIM
J. Bacteriol. 184 (16), 4601-4611 (2002)

REFERENCE
2 (bases 1 to 10393)
AUTHORS Deng,W., Burland,V., Plunkett,G. III, Boutin,A., Mayhew,G.F.,
Liss,P., Perna,N.T., Rose,D.J., Mau,B., Zhou,S., Schwartz,D.C.,
Fetherston,J.D., Lindler,L.E., Brubaker,R.R., Plana,G.V.,
Straley,S.C., McDonough,K.A., Nilles,M.L., Matson,J.S.,
Blattner,F.R. and Perry,R.D.
Genome Sequence of Versinia pestis KIM
J. Bacteriol. 184 (16), 4601-4611 (2002)

Fetherston,J.D., Lindler,L.E., Brubaker,R.R., Plana,G.V.,
Straley,S.C., McDonough,K.A., Nilles,M.L., Matson,J.S.,
Blattner,F.R. and Perry,R.D.
Direct Submission
Submitted (21-FEB-2002) Genetics, University of Wisconsin, 445
Henry Mall, Madison, WI 53706, USA
Location/Qualifiers
1..10393
/organism="Versinia pestis KIM"
/strain="KIM"
/db_xref="taxon:187410"
complement(134..2272)
/gene="nrdB"
/note="y0733"
complement(134..2272)
/gene="nrdB"
/function="enzyme: 2'-Deoxyribonucleotide metabolism"
/note="residues 5 to 712 of 712 are 86.31 pct identical to
residues 5 to 712 of 712 from E. coli K12 : B4238;
residues 5 to 712 of 712 are 87.16 pct identical to
residues 5 to 712 of 712 from GenPept : >gblAAL23272.1]
(AE008908) anaerobic ribonucleoside-triphosphate reductase
[Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="anaerobic ribonucleoside-triphosphate reductase"
/protein_id="AAM84321.1"
/db_xref="GI:21957455"
/translation="MKTVIKRDGCQVPDFEVRIKEAVERAALAVGVDDYCATVAR
VVAQTENQPKVIDREIQNAVENVQMAQYQKLARAYIEVHDRISELRLGNQOEI
QGLVQSNAALLNANKSKVIPTRDLLAGIYAKHYAKQHLPLDVVLAHERGEITH
YHDDYAFPFMFNCMLIDLEMLTQGFMGAIEIPPKSISTATAQIIAIVASHVS
IYGGTTIRIDEVLAPFYTESFNKHKAEEQIPDVAGYAMSRETCYEDAFQSLYE
EVTUHTANGOTPFVTFGGLTSQSMIOQAILNRNIAGLKGRKTATVPKLVFAI
RDGLNHKFGDANYDIKQALECATKRVPDILNVVVKYTVSGTKPMGCRSFVGVE
BEQQIHGCRNGLVISLNPRIALEAKGDNQFWLLDERLLIAKALMTARTARLES
EKARVAPILMEGACGVRLQADDIAEIFKNGRASISLGYIHETHNALFHQTHLFI
DDQLRKRAIAIAVHLKQATDWKDETGYAFSLYSTPENCLDRCDRLFDAEFQVGV
VTDKGYTNSEPHLDVKKRNPYDKLDEAPYPPLANGGFICYGEPNLQHNIRALEVD
WDYSYRVYPYGTNPIDECYECGTFGEFSCSTSGTKGPCKGNHEPSKSVSTRVCY
LGSPDARPNAGKQEEVRKRVKHLANGLG"
complement(2706..3422)
/gene="y0734"
complement(2706..3422)
/gene="y0734"
/function="transport of large molecules: protein, peptide
secretion"
/note="residues 3 to 230 of 238 are 50.64 pct identical to
residues 1 to 226 of 245 from GenPept : >gblAAL45414.1]
(AE009390) ABC transporter, nucleotide binding/ATPase
protein [dipeptide] [Agrobacterium tumefaciens str. C58
(U. Washington)]"
/codon_start=1
/transl_table=11
/product="dipeptide ABC transporter, ATP binding protein"
/protein_id="AAM84322.1"
/db_xref="GI:21957456"
/translation="MALIEVNVLQVSFAQKNQIKTAVESVFAVEANETSLIGASC
CKSTVLRTLAGLOREMOGEITLLGORLPQGORFTQLRKDVQMVFQPYASLHPQHNI
QRTLCEPKINGEKNIQRIQALAEQVGLPASAIRYPHFQSGQRRQRIATRALILLR
PKLLLEDISELSDMSVQAEILNLLNLHLKREYGMTRYLLVSHSDSDVAHMSERAVILME
GHIVREYREDLASAESMSG"
complement(3409..4317)
/gene="dppB"
/note="y0735"
complement(3409..4317)
/gene="dppB"
/function="transport; protein, peptide secretion"
/note="residues 14 to 276 of 302 are 42.06 pct identical
to residues 2 to 258 of 327 from E. coli K12 : B3541;
residues 11 to 295 of 302 are 43.84 pct identical to
residues 1 to 277 of 278 from GenPept : >dbjBAB51953.1]
(AP003006) oligopeptide ABC transporter, ATP-binding

Details of *Y. pestis* sequencing at the Sanger Centre are available on the world wide web.
(URL, http://www.sanger.ac.uk/Projects/Y_pestis/).

FEATURES

```

source
1..216050
/organism="Yersinia pestis"
/strain="CO92"
/db_xref="taxon:632"
/note="biovar: Orientalis"
86..940
/gene="YPO3402"
/note="panC"
86..940
CDS
/gene="YPO3402"
/EC_number="6.3.2.1"
/note="Similar to Escherichia coli pantoate--beta-alanine
ligase PanC SW:PANC_ECOLI (P31663) (283 aa) fasta scores:
E(): 0, 72.0% id in 282 aa, and to Vibrio cholerae
pantoate--beta-alanine ligase Vc0591 TR:Q9KUD1
(EMBL:AE004144) (293 aa) fasta scores: E(): 0, 59.0% id in
283 aa"
/codon_start=1
/transl_table=11
/product="pantoate--beta-alanine ligase"
/protein_id="CAC92632.1"
/db_xref="GI:15981329"
/db_xref="SPTREMBL:Q8ZBK7"
/translation="MLIIETLPLRQQIQRWRQEKRIALVPTMGNLHEGHMTLVDEA
KTRADVVPVITFNPQLQEPDPDLAHYPTLQEDCEKLTIRHGADLVFAPAAADIYPAG
LEKQITVDVVALSTLLEGASRPGHFGVSTIVSKLFLNIQPDVACFGKDKQQALALIR
KWADMGYDINIVGPTVRKDGIALSSRNGVLTEERQIAPOLSKINWALAEKMLG
PROIDALLEEAAQLRVGTFPDELFIQDAETLOPLTVDSQQAVILMAWLGKRLID
NQLVDLRH"
1351..1731
/gene="YPO3403"
/note="panD"
1351..1731
CDS
/gene="YPO3403"
/EC_number="4.1.1.11"
/note="Similar to Escherichia coli aspartate
1-decarboxylase precursor PanD SW:PAND_ECOLI
(P31664) (126
aa) fasta scores: E(): 0, 78.6% id in 126 aa, and to
Pseudomonas aeruginosa aspartate 1-decarboxylase precursor
Pa4731 TR:AAG08117 (EMBL:AE004887) (126 aa) fasta scores:
E(): 1.5e-27, 58.2% id in 122 aa"
/codon_start=1
/transl_table=11
/product="aspartate 1-decarboxylase precursor"
/protein_id="CAC92633.1"
/db_xref="GI:15981330"
/db_xref="SPTREMBL:Q8ZBK6"
/translation="MIRTMLOQKLRHVKTQADLHYEGSCAIDQDFLPAAGILEVEAI
DIYVNDQRFSTYAIAAERGSRIISVNGAAACACVGDGKLIICSYVQMSYAAARLHH
PKVAYFEGENOLQRKAKAVPVQA"
1351..1698
/gene="YPO3403"
/note="Pfam match to entry PF02261 Asp-decarbox, Aspartate
decarboxylase, score 264.30, E-value 1.6e-75"
complement(1880..3073)
/gene="YPO3404"
complement(1880..3073)
/gene="YPO3404"
/note="Similar to Escherichia coli hypothetical protein
YadE SW:YADE_ECOLI (P31666) (409 aa) fasta scores: E(): 0,
65.4% id in 382 aa"
/codon_start=1
/transl_table=11
/product="Conserved hypothetical protein"
/protein_id="CAC92634.1"
/db_xref="GI:15981331"
/db_xref="SPTREMBL:Q8ZBK5"
/translation="MMTERDSEVSLVGEHVPVGEIKEGOLIQVTPVIAEYIEPKFG
NIGTFIDRDLREINKAKQKNDMLGDLNKLPLPNQNILTBREAKVYLAADVRSQFAVL

```

```

AENLRYPITVGKLRSLNTWYQVNIQDRLGYVSADAETDINGIPILTYHHMLKNEENK
RFLNTSTTSDVAFSNOMAYLKQAGYDITISLYQLYLNKNKINLPKVVVLTFFDDGLK
SVSRVAYPILKENGPRATAFIISSRKIKRHPKQNPDSLQFMSIALKLGQDVFDIQLSH
THFLHRTDNRHHPILLRSYHNIIFDPSRRALAQFNPVHIFLSYPPGGFNQTAJYNA
AKNAGFHIAVTTMQGKVRPNDNPYTLKRLYLILRTDSIPTMAERIANQPGPVAVKPAPA
VIEMD"
complement(3400..4170)
/gene="YPO3405"
/note="yadH"
complement(3400..4170)
/gene="YPO3405"
/note="Similar to Escherichia coli hypothetical 28.5 kDa
protein in hpt-panD intergenic region yadH SW:YADH_ECOLI
(P36880) (256 aa) fasta scores: E(): 0, 80.1% id in 256
aa, and to Vibrio cholerae putative permease Vc0590
TR:Q9KUD2 (EMBL:AE004144) (256 aa) fasta scores: E(): 0,
66.0% id in 256 aa"
/codon_start=1
/transl_table=11
/product="putative ABC transporter, integral membrane
protein"
/protein_id="CAC92635.1"
/db_xref="GI:15981332"
/db_xref="SPTREMBL:Q8ZBK4"
/translation="MTRLWIATLQSIWIKEITRFARIWITLVPVITMSLVYVIFGN
LIGARIGDMGGFDYMQFIVPGLIMMAVITNAYSNVASFYGAKFORSIEELLVAPVPT
HVIYIGYGGVGARGICVGLVITIIISLFFVPLHVSWSKIALTLILTAILSGLGLLN
HYAFTFDDISLVPTFFVITPLTYLGGVYSLSLPPFWQAVSKLPIVYMISGFRYGF
LGITDVSLAYTIGVLVFIYFAVYAWAWYLIERTGLRT"
complement(3409..4164)
/gene="YPO3405"
/note="Pfam match to entry PF01061 ABC2_membrane, ABC-2
type transporter, score -22.30, E-value 5.8e-05"
complement(3427..3492)
/gene="YPO3405"
/note="one of 6 probable transmembrane helices predicted
for YPO3405 by TMHMM2.0"
complement(3526..3639)
/gene="YPO3405"
/note="PS00890 ABC-2 type transport system integral
membrane proteins signature."
complement(3592..3657)
/gene="YPO3405"
/note="one of 6 probable transmembrane helices predicted
for YPO3405 by TMHMM2.0"
complement(3679..3744)
/gene="YPO3405"
/note="one of 6 probable transmembrane helices predicted
for YPO3405 by TMHMM2.0"
complement(3790..3855)
/gene="YPO3405"
/note="one of 6 probable transmembrane helices predicted
for YPO3405 by TMHMM2.0"
complement(4045..4110)
/gene="YPO3405"
/note="one of 6 probable transmembrane helices predicted
for YPO3405 by TMHMM2.0"
complement(4167..5093)
/gene="YPO3406"
/note="yadG"
complement(4167..5093)
/gene="YPO3406"
/note="Similar to Escherichia coli hypothetical ABC
transporter ATP-binding protein yadG SW:YADG_ECOLI
(P36879) (308 aa) fasta scores: E(): 0, 81.8% id in 308
aa, and to Vibrio cholerae ABC transporter ATP-binding
protein Vc0589 TR:Q9KUD3 (EMBL:AE004144) (305 aa) fasta
scores: E(): 0, 71.4% id in 304 aa"
/codon_start=1

```

```
/transl_table=11
/product="putative ABC transporter, ATP-binding protein"
/protein_id="CAC92636.1"
/db_xref="GI:1598133"
/db_xref="SPTREMBL:Q8ZBK3"
/translation="MTALEITQTKYMGVQALRGIDLHVEAGDYALLGPNAGK
STTIGISLVNKTSKGKVGQYDIDNDIVNAKROGLVPOEFNFPEVLIQVITQ
AGYGVTRALARAERYLSQDLWSKRDRAIRLSGMRRLMIARALHEPKLLIL
DEPTAGVDIELRRSMWGFLEKLNAGQGTIILTHYLEAEMLCRNIGTIONGELVENT
TMKOLLKLESETIFDLAGKSPKLEGYGYRLDTSTLEVDVYKRGQGLNSLSQLN
VQGVQVOSMRNKARLELEFVTLVNGHEKA"
complement(4458..5003)
/gene="YP03406"
/misc_feature
/note="Pfam match to entry PF00005 ABC_tran, ABC
transporter, score 198.80, E-value 8.5e-56"
complement(4641..4685)
/gene="YP03406"
/misc_feature
/note="PS00211 ABC transporters family signature."
complement(4959..4982)
/gene="YP03406"
/gene="YP00017 ATP/GTP-binding site motif A (P-loop)."
5329..5991
/gene="YP03407"
/note="yadF"
5329..5991
/gene="YP03407"

Query Match 3.1%; Score 21; DB 1; Length 216050;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 GCGTAAATCCATTCCTACCAG 503
|||||
DB 53484 GCGTAAATCCATTCCTACCAG 53464

RESULT 11
AF415022 884 bp DNA linear PLN 10-MAR-2002
DEFINITION Chamaebatiaria millefolium isolate 90Chamaebatiar NADP dependent
sorbitol 6-phosphate dehydrogenase gene, partial cds.
ACCESSION AF415022
KEYWORDS
SOURCE Chamaebatiaria millefolium.
ORGANISM Chamaebatiaria millefolium
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Rosales; Rosaceae; Rosaceae incertae sedis;
Chamaebatiaria.
REFERENCE 1 (bases 1 to 884)
Bortiri,E., Oh,S.-H., Gao,F.-Y. and Potter,D.
Phylogenetic analysis of Prunus (Rosaceae) as determined by
sequences of 56pdh
Unpublished
REFERENCE 2 (bases 1 to 884)
Bortiri,E., Oh,S.-H., Gao,F.-Y. and Potter,D.
Direct Submission
JOURNAL Submitted (29-AUG-2001) Pomology, University of California, Davis,
One Shields Ave, Davis, CA 95616, USA
FEATURES
source
location/Qualifiers
1..884
/organism="Chamaebatiaria millefolium"
/isolate="90Chamaebatiar"
/db_xref="taxon:3222"
join(<223..442,656..>.884)
/product="NADP dependent sorbitol 6-phosphate
dehydrogenase"
join(<223..442,656..>.884)
/codon_start=2
/product="NADP dependent sorbitol 6-phosphate
dehydrogenase"
/protein_id="AAL86687.1"
```

```
/db_xref="GI:19310942"
/translation="NYELFLTRDCLSYAKIKPAVSQFETHPYOYRETIYRFCOKHGVC
PVATPLGGATANTKYGSSPLGDPVLNEVAKYQKVAQICLRWNIERTAVVPKS
TKYERLKENFQVDEKLEKEDIDLINTIDRKSRTTLFSLWSGLDYYA"
BASE COUNT 259 a 171 c 171 g 283 t
ORIGIN
Query Match 3.0%; Score 20; DB 8; Length 884;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 323 GAAAAATTTACTATGTGCAG 342
|||||
DB 203 GAAAAATTTACTATGTGCAG 222

RESULT 12
AB064593/c 7770 bp DNA linear BCT 05-JUL-2001
LOCUS Escherichia coli icdA gene for isocitrate dehydrogenase, complete
DEFINITION Escherichia coli icdA (strain:BL21(DE3)) DNA.
ACCESSION AB064593
VERSION AB064593.1 GI:14624984
KEYWORDS
SOURCE Escherichia coli (strain:BL21(DE3)) DNA.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1
AUTHORS Aoshima,M., Ishii,M., Yamagishi,A., Oshima,T. and Igarashi,Y.
TITLE Sequence polymorphism around the isocitrate dehydrogenase (ICDH)
gene of Escherichia coli BL21(DE3) and citrate accumulation of an
ICDH defective derivative
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 7770)
Aoshima,M.
Direct Submission
JOURNAL Submitted (03-JUL-2001) Miho Aoshima, The University of Tokyo,
Graduate School of Agricultural and Life Sciences; Yayoi 1-1-1,
Bunkyo-ku, Tokyo 113-8657, Japan
(E-mail:aomih@mail.ecc.u-tokyo.ac.jp, Tel:81-3-5841-5143,
Fax:81-3-5841-5272)
FEATURES
source
location/Qualifiers
1..7770
/organism="Escherichia coli"
/strain="BL21(DE3)"
/db_xref="taxon:562"
1138..2388
/gene="icdA"
1138..2388
/gene="icdA"
/codon_start=1
/transl_table=11
/product="isocitrate dehydrogenase"
/protein_id="BAB61874.1"
/db_xref="GI:14624985"
/translation="MESKVVVPAQGGKKTITLQNGKLVNPNPIIYIBSGDIGVDVTPA
MLKVDAAVEKAYKGEKISNMEIYTCESKSTQYQGVWLPAPETLDLIRYRVAIKGP
LTPFVGGGIRSLNVALRQELDLTICLRPRTYOGTSPFVKHPELTDWVIFRENSEDIY
AGIEWKADSADAEEKVIRKFLREMGVKKIRFPEHCIGIGIKPCSEEGTKRLVRAAIEYAI
ANDRDSVTLVHGNIMKFTFEGAKDMGQYOLAREEFGGELIDGGGFWLKKVNPNTKEIV
IKDVHTGTPAKYAGQDKVNPGSIIILSAEMMLRHMGWTEADLIVKMGEGAINAKTVTVYD
FERLMGAKLLKSEFGDAIKNM"
BASE COUNT 2259 a 1574 c 1678 g 2259 t
ORIGIN
Query Match 3.0%; Score 20; DB 1; Length 7770;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 TTCATAAACCCCTACGATGT 155
|||||
```



```
/function="enzyme; purine ribonucleotide biosynthesis"
/notice="residues 4 to 459 are 88.59 pct identical to
residues 1 to 456 of 456 from E. coli K12 : B1131;
residues 4 to 459 are 88.59 pct identical to
residues 1 to 456 of 456 from GenPept :
>gb|AAC55957.1|AE005329.1 (AE005329) adenylsuccinate
lyase [Escherichia coli O157:H7 EDL933]"
/codon_start=1
/transl_table=1
/product="adenylsuccinate lyase"
/protein_id="AAM85365.1"
/db_xref="GI:21958606"
/transl_table="MPPMELSSLTAVSPIDRGYDKVSALRPISFEGLLKFRVQVEV
RWQKLAACIEKVEPAPFADANANAYLDKIYEFNEQDAQRIKTIERTNHDVRAVEYF
LKKEVNEPALHAYSEIHFACSTEDINNLSHALMLQTAQDVLPAWROIIDSIKAL
AHQYRDLPPLSRTHGQPTSTIGKELANVAYRMERFROLTOVEIIGKNGAVGYN
AHIVAYPEVDWHPQSESVTSLGNNWNPYTOLEPHDYIAELFDCVARENTLIDFR
DINGYIALNFKQKTIAGEIGSSTMPHKVNPIDFENSEGNLGLSNVGHMKSLPVS
RWQDRDTSVLRNIGLVGLYALIAIYQATMKGISKLEVNHAHLIQELDHNEVLAEP
QTIMRRYGIKPYEKLKELTRGRKVDAAQMAQAFIDGLDLPPEEKRLKAMTPANYIGR
ATTMVDELK"
complement(5936. .6568)
/gene="y1798"
complement(5936. .6568)
/gene="y1798"
/notice="residues 3 to 209 of 210 are 75.48 pct identical to
residues 1 to 208 of 213 from E. coli K12 : B1132"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAM85366.1"
/db_xref="GI:21958607"
/transl_table="MIVAKNYDITLALAGICQSARLVQOLAHEGQDNDALNTVLRG
LLQTNPSSTLAVGTDTQVLKMGLETQSVLNANROGAEALRYTLSLMVLGKLSA
SKSAMNTLGRISQDLROLAHFDESETMSSLASIYDVVSPGPRIOVTGSPAILQ
SPLQAKVRLTLGAGIRSAVLWQVGSRLQLMFSRNLKFKQAQSIILAH"
complement(6620. .7735)
/gene="y1799"
complement(6620. .7735)
/gene="y1799"
/notice="residues 7 to 367 of 371 are 87.53 pct identical to
residues 21 to 381 of 383 from E. coli K12 : B1133;
residues 1 to 371 of 371 are 100.00 pct identical to
residues 1 to 371 of 371 from GenPept : >emb|CAC90460.1|
(AJ414149) tRNA
(5-methylaminomethyl-2-thiouridylate)-methyltransferase
[versinia pestis]"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAM85367.1"
/db_xref="GI:21958608"
/transl_table="MSDNSQKVVIGMVGSDVSVSAYLLQQQYQVAGLFMKNWEED
DDEYCSAATDLADAQVCKLGMELHTVNFRAEYNDVNFELFAETKAGRTPNPDIL
KNEIKFALEFAEEDLGADYIATGHYVRQVDGKSRLLGLDGNKQDSYFLYLS
HQEIOASLPVGELEKPEVRIAEQLDLVATKKKDSGTGICFIFGERFRDPLGRYLPQ
PGPTMTVDGLVGHQGLMYHTLQGRGLIGTGKEGDDPWVWPKDLDNSNTLLVAQ
GHEHPRLMSVLGVAOOLLHWVDROPVTPAPRCVVKTRYRQQDICTVTPLDDEVDYRF
DDPVAAVTPQSAVFYQGEICLGGIIEQRYPLTNP"
```

Query Match 3.0%; Score 20; DB 1; Length 10473;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 520 CTTTATGAAGCGCGCAACCG 539
|||||

Db 8504 CTTTATGAAGCGCGCAACCG 8485
|||||

RESULT 14

U55370

LOCUS

DEFINITION

U55370 32403 bp DNA linear INV 23-MAY-2002
Caenorhabditis elegans cosmid K03B4, complete sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

U55370
U55370.1 GI:1280096
HTG.
Caenorhabditis elegans.
Caenorhabditis elegans

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
Waterston,R.

Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium
Science 282 (5396), 2012-2018 (1998)
99069613
9851916

REFERENCE
1 (bases 1 to 32403)

2 (bases 1 to 32403)

Du,Z., Le,T.T. and Kemp,K.

The sequence of C. elegans cosmid K03B4

Unpublished (2001)

REFERENCE 3 (bases 1 to 32403)

Waterston,R.

Direct Submission

Submitted (17-APR-1996) Robert Waterston

4 (bases 1 to 32403)

Waterston,R.

Direct Submission

Submitted (01-AUG-2001) Department of Genetics, Washington

University, Genome Sequencing Center, 4444 Forest Park Avenue, St.

Louis, MO 63110, USA

5 (bases 1 to 32403)

Waterston,R.

Direct Submission

Submitted (01-JAN-2002) Department of Genetics, Washington

University, Genome Sequencing Center, 4444 Forest Park Avenue, St.

Louis, MO 63110, USA

6 (bases 1 to 32403)

Waterston,R.

Direct Submission

Submitted (23-MAY-2002) Department of Genetics, Washington

University, Genome Sequencing Center, 4444 Forest Park Avenue, St.

Louis, MO 63110, USA

Submitted by:

Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
email: rwenematode.wustl.edu and jess@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate chemistry
or covered by high quality data (i.e., phred quality >= 30); an
attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by sequence from
more than one ml3 subclone.

For a graphical representation of this cosmid sequence and its
analysis see:

<http://www.wormbase.org/db/seq/sequence?name=K03B4;class=Sequence>

NEIGHBORING COSMID INFORMATION

The 5' cosmid is F41F3, 200 bp overlap; the 3' cosmid is ZC266,
3000 bp overlap. Actual start of this cosmid is at base position
197 of K03B4; actual end is at 5381 of ZC266.

NOTES:

Coding sequences below are the result of integration and manual

review of the following data : computer analysis using the program Genefinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yuji Kohara (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C. elegans ORFome cloning project (http://worldb.dfci.harvard.edu/), similarity to other proteins from BlastX analyses (http://blast.wustl.edu/), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans Genbank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES
source

Location/Qualifiers
1. .32403
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="v"
/clone="K03B4"
complement(2043. .3732)
/gene="K03B4.5"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=K03B4.5;class
=Sequence"
complement(join(2043. .2121,2169. .2442,2631. .2863,
3054. .3214,3258. .3340,3558. .3732))
/gene="K03B4.5"
/note="weak similarity to Xenopus dopamine D2 receptor
(SP:P24628)"
/codon_start=1
/product="Hypothetical protein K03B4.5"
/protein_id="AA97993.1"
/db_xref="GI:1280097"
/translation="MGHSDTRFRKRVDPFAPTFVNFNPLKISSENLIIPILISVY
LNFQVFQEDSICHTMTSTFTTISDWNAAAALVMTNCTGFFNFLLITFS
FYDKQKTSFLNLCVFRSINNVAILFILVYIPALIIIGESIYHPMVEVLTITAMNL
KYVNEFQSYLIRNLVAIYFPLKYNLFGILKLAHFIIYLDVRNVVTFENIDRYK
DSKFMFLSVKHLAYGCVFVDPGIFWSGLIIFPFVNIETFAIRLYLKRKTQNA
N"NKDIKKNALFQIIFQDSLFFISVAFMTKMNMLIDHRYFFSFFSQTFLWQSIHVDG
complement(5626. .10741)
/gene="K03B4.6"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=K03B4.6;class
=Sequence"
complement(join(5626. .5916,5967. .6269,6725. .6901,
8300. .8587,8644. .8937,8982. .9150,10475. .10593,
10697. .10741))
/gene="K03B4.6"
/codon_start=1
/product="Hypothetical protein K03B4.6"
/protein_id="AA97994.1"
/db_xref="GI:1280098"

/translation="MNGEYFLEQAHLISFNATLGATSDPFFLNTSAAGNPTKGTENP
PYVEGFMAYILFRLSMILFQKTVFSLFAFATSLIEEDVRDVCYKWDIAKTPDG
CIKRIKYVPDVGVEIGTKFCSEISKVCRILKTDTEKNYNTFELDDITPKITANESDC
YDPNVEFQCRVLQGRNALOMLVDPYENKNVDKHAHAKISKQCOIVKNIJDTCTPE
DSVDFIVEGCVINLKQTSFMDCLSKLTKKTPNLSKYTCENFQTKNITLKRIPT
ADKMCWELMRKHGESSVDFQKNAATNCYSIKQSEVKEFKESKRLDCKDQDFVI
GGVPCIEFFKFIYQVYLTIKSNDQDQVHSHKHLCDLIDDEYNKEYERLWGDQKV
EDTGLFMIOGLRELTECESCHNSGIEIGNLKEILAMEIYGRNYSWKLYDIRRP
SMKTLMEKLLCLRDSDQECRRKFVEKICGNMGNTGTETPWVTEYTTDD"
18524. .19556
/gene="K03B4.4"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=K03B4.4;class
=Sequence"
join(18524. .18578,19300. .19472,19521. .19556)
/gene="K03B4.4"
/codon_start=1
/product="Hypothetical protein K03B4.4"
/protein_id="AA97995.1"

gene

complement(5626. .10741)
/gene="K03B4.6"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=K03B4.6;class
=Sequence"
complement(join(5626. .5916,5967. .6269,6725. .6901,
8300. .8587,8644. .8937,8982. .9150,10475. .10593,
10697. .10741))
/gene="K03B4.6"
/codon_start=1
/product="Hypothetical protein K03B4.6"
/protein_id="AA97994.1"
/db_xref="GI:1280098"

CDS

/translation="MNGEYFLEQAHLISFNATLGATSDPFFLNTSAAGNPTKGTENP
PYVEGFMAYILFRLSMILFQKTVFSLFAFATSLIEEDVRDVCYKWDIAKTPDG
CIKRIKYVPDVGVEIGTKFCSEISKVCRILKTDTEKNYNTFELDDITPKITANESDC
YDPNVEFQCRVLQGRNALOMLVDPYENKNVDKHAHAKISKQCOIVKNIJDTCTPE
DSVDFIVEGCVINLKQTSFMDCLSKLTKKTPNLSKYTCENFQTKNITLKRIPT
ADKMCWELMRKHGESSVDFQKNAATNCYSIKQSEVKEFKESKRLDCKDQDFVI
GGVPCIEFFKFIYQVYLTIKSNDQDQVHSHKHLCDLIDDEYNKEYERLWGDQKV
EDTGLFMIOGLRELTECESCHNSGIEIGNLKEILAMEIYGRNYSWKLYDIRRP
SMKTLMEKLLCLRDSDQECRRKFVEKICGNMGNTGTETPWVTEYTTDD"
18524. .19556
/gene="K03B4.4"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=K03B4.4;class
=Sequence"
join(18524. .18578,19300. .19472,19521. .19556)
/gene="K03B4.4"
/codon_start=1
/product="Hypothetical protein K03B4.4"
/protein_id="AA97995.1"

gene

complement(5626. .10741)
/gene="K03B4.6"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=K03B4.6;class
=Sequence"
complement(join(5626. .5916,5967. .6269,6725. .6901,
8300. .8587,8644. .8937,8982. .9150,10475. .10593,
10697. .10741))
/gene="K03B4.6"
/codon_start=1
/product="Hypothetical protein K03B4.6"
/protein_id="AA97994.1"
/db_xref="GI:1280098"

CDS

/translation="MNGEYFLEQAHLISFNATLGATSDPFFLNTSAAGNPTKGTENP
PYVEGFMAYILFRLSMILFQKTVFSLFAFATSLIEEDVRDVCYKWDIAKTPDG
CIKRIKYVPDVGVEIGTKFCSEISKVCRILKTDTEKNYNTFELDDITPKITANESDC
YDPNVEFQCRVLQGRNALOMLVDPYENKNVDKHAHAKISKQCOIVKNIJDTCTPE
DSVDFIVEGCVINLKQTSFMDCLSKLTKKTPNLSKYTCENFQTKNITLKRIPT
ADKMCWELMRKHGESSVDFQKNAATNCYSIKQSEVKEFKESKRLDCKDQDFVI
GGVPCIEFFKFIYQVYLTIKSNDQDQVHSHKHLCDLIDDEYNKEYERLWGDQKV
EDTGLFMIOGLRELTECESCHNSGIEIGNLKEILAMEIYGRNYSWKLYDIRRP
SMKTLMEKLLCLRDSDQECRRKFVEKICGNMGNTGTETPWVTEYTTDD"
18524. .19556
/gene="K03B4.4"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=K03B4.4;class
=Sequence"
join(18524. .18578,19300. .19472,19521. .19556)
/gene="K03B4.4"
/codon_start=1
/product="Hypothetical protein K03B4.4"
/protein_id="AA97995.1"

gene

/db_xref="GI:1280099"
/translation="MRFLAAALLLVAVLLATFDVTSQSRIDRSAAWFKPLPNVRSPQ
YRSRIGSGSEILLPRMYRSLPQNDYQGMKDLITDQMDAEN"
21710. .22411
/gene="K03B4.3"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=K03B4.3a;clas
s=Sequence"
join(21710. .21784,21827. .21894,21959. .22223,22280. .22411)
/gene="K03B4.3"
/note="coded for by the following C. elegans cDNAs:
yk602f4.3, yk602f4.5"
/codon_start=1
/product="Hypothetical protein K03B4.3b"
/protein_id="AAK82906.1"
/db_xref="GI:15055382"

CDS

/translation="MNDPEQYEPSSSTESVLMPPALPOYFORPAAAPQVYSTLEPSV
QNLRSLSHLKPLTGNOQFQVKLEQKNPQSDDTHEFINOLADYPTTPDSTVTLHFL
KSAGVDGSDPRVTRMISLAAQKHVSDIILDAMTSARMKGLGOTKGTGKDTKYLTEEL
LDEILKEGHQNTRPYHT"
join(21710. .21784,21836. .21894,21959. .22223,22280. .22411)
/gene="K03B4.3"
/note="similar to human TATA-binding protein associated
factor 30 kDa subunit (PIR:A54981); coded for by the
following C. elegans cDNAs: yk70h11.5, yk331g8.3,
yk331g8.5"
/codon_start=1
/product="Hypothetical protein K03B4.3a"
/protein_id="AA97996.1"
/db_xref="GI:1280100"

CDS

/translation="MNDPEQYEPSSSTESVLMPPALPOYFORPAAAPQVYSTLEPSVONL
RSLHLKPLTGNOQFQVKLEQKNPQSDDTHEFINOLADYPTTPDSTVTLHFLKSA
YVDGSDPRVTRMISLAAQKHVSDIILDAMTSARMKGLGOTKGTGKDTKYLTEELDE
ILKEGHQNTRPYHT"
22603. .23470
/gene="K03B4.2"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=K03B4.2;class
=Sequence"
join(22603. .22837,23253. .23470)
/gene="K03B4.2"
/note="coded for by the following C. elegans cDNAs:
yk99d5.5, yk113d4.5, yk117d8.5, yk158b11.5, yk172h12.5,
yk113d4.3, yk99d5.3, yk117d8.3, yk158b11.3, yk172h12.3"
/codon_start=1
/product="Hypothetical protein K03B4.2"
/protein_id="AA97997.1"
/db_xref="GI:1280101"

gene

/translation="MNDPEQYEPSSSTESVLMPPALPOYFORPAAAPQVYSTLEPSVONL
RSLHLKPLTGNOQFQVKLEQKNPQSDDTHEFINOLADYPTTPDSTVTLHFLKSA
YVDGSDPRVTRMISLAAQKHVSDIILDAMTSARMKGLGOTKGTGKDTKYLTEELDE
ILKEGHQNTRPYHT"
22603. .23470
/gene="K03B4.2"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=K03B4.2;class
=Sequence"
join(22603. .22837,23253. .23470)
/gene="K03B4.2"
/note="coded for by the following C. elegans cDNAs:
yk99d5.5, yk113d4.5, yk117d8.5, yk158b11.5, yk172h12.5,
yk113d4.3, yk99d5.3, yk117d8.3, yk158b11.3, yk172h12.3"
/codon_start=1
/product="Hypothetical protein K03B4.2"
/protein_id="AA97997.1"
/db_xref="GI:1280101"

CDS

/translation="MNDPEQYEPSSSTESVLMPPALPOYFORPAAAPQVYSTLEPSVONL
RSLHLKPLTGNOQFQVKLEQKNPQSDDTHEFINOLADYPTTPDSTVTLHFLKSA
YVDGSDPRVTRMISLAAQKHVSDIILDAMTSARMKGLGOTKGTGKDTKYLTEELDE
ILKEGHQNTRPYHT"
22603. .23470
/gene="K03B4.2"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=K03B4.2;class
=Sequence"
join(22603. .22837,23253. .23470)
/gene="K03B4.2"
/note="coded for by the following C. elegans cDNAs:
yk99d5.5, yk113d4.5, yk117d8.5, yk158b11.5, yk172h12.5,
yk113d4.3, yk99d5.3, yk117d8.3, yk158b11.3, yk172h12.3"
/codon_start=1
/product="Hypothetical protein K03B4.2"
/protein_id="AA97997.1"
/db_xref="GI:1280101"

gene

/translation="MNDPEQYEPSSSTESVLMPPALPOYFORPAAAPQVYSTLEPSVONL
RSLHLKPLTGNOQFQVKLEQKNPQSDDTHEFINOLADYPTTPDSTVTLHFLKSA
YVDGSDPRVTRMISLAAQKHVSDIILDAMTSARMKGLGOTKGTGKDTKYLTEELDE
ILKEGHQNTRPYHT"
22603. .23470
/gene="K03B4.2"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=K03B4.2;class
=Sequence"
join(22603. .22837,23253. .23470)
/gene="K03B4.2"
/note="coded for by the following C. elegans cDNAs:
yk99d5.5, yk113d4.5, yk117d8.5, yk158b11.5, yk172h12.5,
yk113d4.3, yk99d5.3, yk117d8.3, yk158b11.3, yk172h12.3"
/codon_start=1
/product="Hypothetical protein K03B4.2"
/protein_id="AA97997.1"
/db_xref="GI:1280101"

gene

/translation="MNDPEQYEPSSSTESVLMPPALPOYFORPAAAPQVYSTLEPSVONL
RSLHLKPLTGNOQFQVKLEQKNPQSDDTHEFINOLADYPTTPDSTVTLHFLKSA
YVDGSDPRVTRMISLAAQKHVSDIILDAMTSARMKGLGOTKGTGKDTKYLTEELDE
ILKEGHQNTRPYHT"
22603. .23470
/gene="K03B4.2"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=K03B4.2;class
=Sequence"
join(22603. .22837,23253. .23470)
/gene="K03B4.2"
/note="coded for by the following C. elegans cDNAs:
yk99d5.5, yk113d4.5, yk117d8.5, yk158b11.5, yk172h12.5,
yk113d4.3, yk99d5.3, yk117d8.3, yk158b11.3, yk172h12.3"
/codon_start=1
/product="Hypothetical protein K03B4.2"
/protein_id="AA97997.1"
/db_xref="GI:1280101"

gene

/translation="MNDPEQYEPSSSTESVLMPPALPOYFORPAAAPQVYSTLEPSVONL
RSLHLKPLTGNOQFQVKLEQKNPQSDDTHEFINOLADYPTTPDSTVTLHFLKSA
YVDGSDPRVTRMISLAAQKHVSDIILDAMTSARMKGLGOTKGTGKDTKYLTEELDE
ILKEGHQNTRPYHT"
22603. .23470
/gene="K03B4.2"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=K03B4.2;class
=Sequence"
join(22603. .22837,23253. .23470)
/gene="K03B4.2"
/note="coded for by the following C. elegans cDNAs:
yk99d5.5, yk113d4.5, yk117d8.5, yk158b11.5, yk172h12.5,
yk113d4.3, yk99d5.3, yk117d8.3, yk158b11.3, yk172h12.3"
/codon_start=1
/product="Hypothetical protein K03B4.2"
/protein_id="AA97997.1"
/db_xref="GI:1280101"

gene

/translation="MNDPEQYEPSSSTESVLMPPALPOYFORPAAAPQVYSTLEPSVONL
RSLHLKPLTGNOQFQVKLEQKNPQSDDTHEFINOLADYPTTPDSTVTLHFLKSA
YVDGSDPRVTRMISLAAQKHVSDIILDAMTSARMKGLGOTKGTGKDTKYLTEELDE
ILKEGHQNTRPYHT"
22603. .23470
/gene="K03B4.2"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=K03B4.2;class
=Sequence"
join(22603. .22837,23253. .23470)
/gene="K03B4.2"
/note="coded for by the following C. elegans cDNAs:
yk99d5.5, yk113d4.5, yk117d8.5, yk158b11.5, yk172h12.5,
yk113d4.3, yk99d5.3, yk117d8.3, yk158b11.3, yk172h12.3"
/codon_start=1
/product="Hypothetical protein K03B4.2"
/protein_id="AA97997.1"
/db_xref="GI:1280101"

gene

/translation="MNDPEQYEPSSSTESVLMPPALPOYFORPAAAPQVYSTLEPSVONL
RSLHLKPLTGNOQFQVKLEQKNPQSDDTHEFINOLADYPTTPDSTVTLHFLKSA
YVDGSDPRVTRMISLAAQKHVSDIILDAMTSARMKGLGOTKGTGKDTKYLTEELDE
ILKEGHQNTRPYHT"
22603. .23470
/gene="K03B4.2"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=K03B4.2;class
=Sequence"
join(22603. .22837,23253. .23470)
/gene="K03B4.2"
/note="coded for by the following C. elegans cDNAs:
yk99d5.5, yk113d4.5, yk117d8.5, yk158b11.5, yk172h12.5,
yk113d4.3, yk99d5.3, yk117d8.3, yk158b11.3, yk172h12.3"
/codon_start=1
/product="Hypothetical protein K03B4.2"
/protein_id="AA97997.1"
/db_xref="GI:1280101"

gene

/translation="MNDPEQYEPSSSTESVLMPPALPOYFORPAAAPQVYSTLEPSVONL
RSLHLKPLTGNOQFQVKLEQKNPQSDDTHEFINOLADYPTTPDSTVTLHFLKSA
YVDGSDPRVTRMISLAAQKHVSDIILDAMTSARMKGLGOTKGTGKDTKYLTEELDE
ILKEGHQNTRPYHT"
22603. .23470
/gene="K03B4.2"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=K03B4.2;class
=Sequence"
join(22603. .22837,23253. .23470)
/gene="K03B4.2"
/note="coded for by the following C. elegans cDNAs:
yk99d5.5, yk113d4.5, yk117d8.5, yk158b11.5, yk172h12.5,
yk113d4.3, yk99d5.3, yk117d8.3, yk158b11.3, yk172h12.3"
/codon_start=1
/product="Hypothetical protein K03B4.2"
/protein_id="AA97997.1"
/db_xref="GI:1280101"

gene

/translation="MNDPEQYEPSSSTESVLMPPALPOYFORPAAAPQVYSTLEPSVONL
RSLHLKPLTGNOQFQVKLEQKNPQSDDTHEFINOLADYPTTPDSTVTLHFLKSA
YVDGSDPRVTRMISLAAQKHVSDIILDAMTSARMKGLGOTKGTGKDTKYLTEELDE
ILKEGHQNTRPYHT"
22603. .23470
/gene="K03B4.2"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=K03B4.2;class
=Sequence"
join(22603. .22837,23253. .23470)
/gene="K03B4.2"
/note="coded for by the following C. elegans cDNAs:
yk99d5.5, yk113d4.5, yk117d8.5, yk158b11.5, yk172h12.5,
yk113d4.3, yk99d5.3, yk117d8.3, yk158b11.3, yk172h12.3"
/codon_start=1
/product="Hypothetical protein K03B4.2"
/protein_id="AA97997.1"
/db_xref="GI:1280101"

Query Match 3.0%; Score 20; DB 3; Length 32403;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 CGAAACTCTTTTAGAAAA 60
|||||
DB 12095 CGAAACTCTTTTAGAAAA 12114

RESULT 15
AC106775/c
LOCUS
AC106775 Homo sapiens chromosome 5 clone RP11-322k12, complete sequence.
DEFINITION
AC106775
ACCESSION
AC106775.3 GI:19848351
VERSION
HTG.
KEYWORDS
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

* 2897 4278: contig of 1382 bp in length
* 4279 gap of unknown length
* 4379 contig of 1249 bp in length
* 5627 gap of unknown length
* 5628 7521: contig of 1794 bp in length
* 5728 7621: gap of unknown length
* 7622 8878: contig of 1257 bp in length
* 8879 gap of unknown length
* 8979 11131: contig of 2153 bp in length
* 11132 11231: gap of unknown length
* 11232 14216: contig of 2985 bp in length
* 14317 14316: gap of unknown length
* 14317 16088: contig of 1772 bp in length
* 16089 16188: gap of unknown length
* 16189 17922: contig of 1734 bp in length
* 17923 18022: gap of unknown length
* 18023 20155: contig of 2133 bp in length
* 20156 20255: gap of unknown length
* 20256 23371: contig of 3116 bp in length
* 23372 23471: gap of unknown length
* 23472 26665: contig of 3194 bp in length
* 26666 26765: gap of unknown length
* 26766 28982: contig of 2217 bp in length
* 28983 29082: gap of unknown length
* 29083 32399: contig of 3317 bp in length
* 32400 32499: gap of unknown length
* 32500 35642: contig of 3143 bp in length
* 35643 35742: gap of unknown length
* 35743 38529: contig of 2787 bp in length
* 38530 38629: gap of unknown length
* 38630 41401: contig of 2772 bp in length
* 41402 41501: gap of unknown length
* 41502 47614: contig of 6013 bp in length
* 47615 50863: contig of 3249 bp in length
* 50864 50963: gap of unknown length
* 50964 56189: contig of 5226 bp in length
* 56190 56289: gap of unknown length
* 56290 61081: contig of 4792 bp in length
* 61082 61181: gap of unknown length
* 61182 68596: contig of 7415 bp in length
* 68597 68696: gap of unknown length
* 68697 75882: contig of 7186 bp in length
* 75883 75982: gap of unknown length
* 75983 80902: contig of 4920 bp in length
* 80903 81002: gap of unknown length
* 81003 88297: contig of 7294 bp in length
* 88297 88397: gap of unknown length
* 88397 96759: contig of 8363 bp in length
* 96760 96859: gap of unknown length
* 96860 103626: contig of 6767 bp in length
* 103627 103726: gap of unknown length
* 103727 111301: contig of 7575 bp in length
* 111302 111401: gap of unknown length
* 111402 123190: contig of 11789 bp in length
* 123191 123291: gap of unknown length
* 123291 152037: contig of 28747 bp in length.

FEATURES
source

1..152037
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-347L16"
BASE COUNT 41109 a 30445 c 29440 g 43496 t 7547 others
ORIGIN

Query Match 3.0%; Score 20; DB 2; Length 152037;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 TGAATAATACCATGACGAAAA 46
|||||
Db 39182 TGAATAATACCATGACGAAAA 39163

RESULT 17
AC004021
LOCUS
DEFINITION Human PAC clone RP1-186K10 from 5q31, complete sequence.
ACCESSION AC004021
VERSION AC004021.1 GI:2795824
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 171370)
AUTHORS Kalicki,J., Kellen,J. and O'Brien,D.
TITLE The sequence of H. sapiens PAC clone RP1-186K10
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 171370)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE
3 (bases 1 to 171370)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE
4 (bases 1 to 171370)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (04-FEB-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics

Center project name: H_DJ0186K10

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

This clone was provided for sequencing by Dr. Michelle M. Le Beau in the Department of Medicine, University of Chicago, Chicago IL; and Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this clone, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-1, prepared by Pieter de Jong and coworkers at Roswell Park Cancer Institute, using the method described by Ioannou et al., Nature Genetics, 6:84-9 (1994). The library is from one male donor. For further details, see http://bacpac.med.buffalo.edu/
The clone is available from Genome Systems, Inc.
(http://www.genomesystems.com).

VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

The actual start of this clone is at base position 1 of RP1-186K10; actual end is at 171370 of RP1-186K10. The orientation of this clone is unknown.

This clone contains STS APM350yb1 (NID:g1051891) and EST53547 (NID:g1347444).

FEATURES

Source

Location/Qualifiers

1. 171370

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="5"

/map="5q31"

/clone="RP1-186K10"

/clone_lib="RPC1-1"

1. 11157

/rpt_family="L1"

1290. 2612

/rpt_family="L1"

2617. 2679

/rpt_family="MER81"

3511. 3643

/rpt_family="MER1_type"

3791. 4239

/rpt_family="L2"

4249. 4551

/rpt_family="MaLR"

4553. 4936

/rpt_family="Retroviral"

5017. 5328

/rpt_family="MER2_type"

5361. 5408

/rpt_family="Retroviral"

5409. 5542

/rpt_family="MaLR"

5546. 5893

/rpt_family="L2"

5920. 5994

/rpt_family="MaLR"

6030. 6172

/rpt_family="L2"

6193. 6236

/rpt_family="L2"

6299. 6504

/rpt_family="L2"

6505. 6536

/rpt_family="(TGAA)n"

7620. 8444

/rpt_family="L1"

8919. 8998

/rpt_family="Retroviral"

9041. 10574

/note="CpG_island (GC=73.7, o/e=0.70, #CpGs=115)"

10335. 10420

/rpt_family="Alu"

10416. 10523

/rpt_family="Alu"

10535. 10594

/rpt_family="Alu"

10732. 10988

/rpt_family="L1"

11023. 14892

/rpt_family="L1"

15585. 15760

/rpt_family="MER1_type"

18900. 18902

/rpt_family="MIR"

18903. 18949

/rpt_family="L2"

18950. 18972

/rpt_family="MIR"

19618. 19768

/rpt_family="L1"

19769. 20895

/rpt_family="L1"

20863. 22033

/rpt_family="L1"

22147. 24390

repeat_region /rpt_family="L1"
24614. 25397
/rpt_family="L1"
25396. 26577
/rpt_family="L1"
27204. 27824
/rpt_family="L1"
28298. 28456
/rpt_family="Alu"
28460. 28510
/rpt_family="7SLRNA"
28548. 28834
/rpt_family="Alu"
28855. 29167
/rpt_family="Alu"
29768. 30055
/rpt_family="L1"
31184. 31477
/rpt_family="Alu"
32124. 32489
/rpt_family="MaLR"
32847. 32733
/rpt_family="MIR"
33532. 33594
/rpt_family="L2"
33613. 33706
/rpt_family="L1"
34315. 34609
/rpt_family="Alu"
34782. 34932
/rpt_family="L2"
34935. 35152
/rpt_family="MIR"
36512. 36802
/rpt_family="Alu"
36882. 37126
/rpt_family="L2"
37957. 38046
/rpt_family="L2"
38742. 38863
/rpt_family="MIR"
40308. 40599
/rpt_family="Alu"
41127. 41200
/rpt_family="MaLR"
41356. 41728
/rpt_family="MaLR"
41917. 42152
/rpt_family="L2"
42169. 42459
/rpt_family="Alu"
42468. 42623
/rpt_family="L2"
43252. 43548
/rpt_family="Alu"
43535. 43594
/rpt_family="(TAA)n"

Query Match

Best Local Similarity 3.0%; Score 20; DB 9; Length 171370;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 541 CAGGTCAGGCGCATGACTGC 560

Db 110782 CAGGTCAGGCGCATGACTGC 110801

RESULT 18

AC116356/c

LOCUS AC116356

DEFINITION Homo sapiens chromosome 5 clone RP11-654L21, WORKING DRAFT

SEQUENCE, 3 unordered pieces.

ACCESSION AC116356

175838 bp DNA linear HTG 10-APR-2002

VERSION	AC116356.2	GI:20127977	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS	Homo sapiens.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 175838)		
AUTHORS	DOE Joint Genome Institute.		
TITLE	Sequencing of Human Chromosome 5		
REFERENCE	2 (bases 1 to 175838)		
AUTHORS	DOE Joint Genome Institute.		
TITLE	Direct Submission		
JOURNAL	Submitted (27-MAR-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA		
REFERENCE	3 (bases 1 to 175838)		
AUTHORS	DOE Joint Genome Institute.		
TITLE	Direct Submission		
JOURNAL	Submitted (10-APR-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA		
COMMENT	On Apr 10, 2002 this sequence version replaced gi:19745038. -----Genome Center Center: Joint Genome Institute Center Code: JGI Web site: http://www.jgi.doe.gov ----- Project Information Center Project Name: 1545925 Center clone name: RPCI-11_654L21 ----- Summary Statistics Consensus quality: 173992 bases at least Q40 Consensus quality: 174716 bases at least Q30 Consensus quality: 175005 bases at least Q20 Estimated insert size: 175000; agarose-fp estimation Estimated insert size: 175638; sum-of-contigs estimation Quality coverage: 5.82 in Q20 bases; agarose-fp estimation Quality coverage: 5.79 in Q20 bases; sum-of-contigs estimation. * NOTE: This is a 'working draft' sequence. It currently * consists of 3 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * * 1 21920: contig of 21920 bp in length * 21921 22020: gap of unknown length * 22021 53261: contig of 31241 bp in length * 53262 53361: gap of unknown length * 53362 175838: contig of 122477 bp in length. Location/Qualifiers 1. .175838 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="5" /clone="RP11-654L21" /clone.lib="RPCI human BAC library 11" 47733 a 36757 c 39091 g 52038 t 219 others		
BASE COUNT	47733 a 36757 c 39091 g 52038 t 219 others		
ORIGIN	Query Match 3.0%; Score 20; DB 2; Length 175838; Best Local Similarity 100.0%; Pred. No. 15; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0; Qy 541 CAGGTCAGCGCATGACTGC 560 Db 74219 CAGGTCAGCGCATGACTGC 74200 RESULT 19 AJ414149/c LOCUS		
DEFINITION	Yersinia pestis strain CO92 complete genome; segment 9/20.		
ACCESSION	AJ414149 AL590842		
VERSION	AJ414149.1 GI:15979570		
KEYWORDS	Yersinia pestis.		
SOURCE	Yersinia pestis		
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Yersinia.		
REFERENCE	1 (bases 1 to 193050)		
AUTHORS	Parkhill, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T.G., Prentice, M.B., Sebaihia, M., James, K.D., Churcher, C., Mungall, K.L., Baker, S., Basham, D., Bentley, S.D., Brooks, K., Cerdeno-Tarraga, A.M., Chillingworth, T., Cronin, A., Davies, R.M., Davis, P., Dougan, G., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Leather, S., Karlyshev, A.V., Moule, S., Oyston, P.C.F., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S. and Barrell, B.G.		
TITLE	Genome sequence of Yersinia pestis, the causative agent of plague		
JOURNAL	Nature 413 (6855), 523-527 (2001)		
MEDLINE	21470413		
PUBMED	11586360		
REFERENCE	2 (bases 1 to 193050)		
AUTHORS	Parkhill, J.		
TITLE	Direct Submission		
JOURNAL	Submitted (04-OCT-2001) Submitted on behalf of the Yersinia sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk		
COMMENT	Notes: Details of Y. pestis sequencing at the Sanger Centre are available on the World Wide Web. (URL, http://www.sanger.ac.uk/projects/Y_pestis/).		
FEATURES	source 1. .193050 /organism="Yersinia pestis" /strain="CO92" /db_xref="taxon:632" /note="biovar: Orientalis" 372. .1151 /gene="YpO1528" 372. .1151 /gene="YpO1528" /note="Similar to Escherichia coli ferric iron reductase protein FhuF SW:PHUF_ECOLI (P39405) (262 aa) fasta scores: E(): 8.4e-33, 35.0% id in 246 aa" /codon_start=1 /transl_table=11 /product="putative ferric iron reductase" /protein_id="CAC90351.1" /db_xref="GI:15979571" /db_xref="SPTREMBL:O8ZF25" /translation="MSNTAQLITPSPGLIADITELFEKTPFAHFSRTLKVNADDIPEE TMSFHTWSSIDNFFTLQKYRDEYGYGNDLKPNDKALYSLWSQYFGLIIPMMLLLI EYPOTIDTHHKFKVLFPHSGRPEVYIQLKWSQDPTGLLERLYLLNHHVPIAEK IESYQINGRLLNWNIIGYLMFVGLGFEKRGDLYQSYINGLFLMELSLPNGQDNPLY RTVMRLNGTLQRRSCCQRNKLPGVRSCHDCPLKPLNLIS" 1303. .2850 /gene="YpO1529" 1303. .2850 /gene="YpO1529" /note="Similar to Enterobacter aerogenes L-2,4-diaminobutyrate decarboxylase ddc TR:Q9S0P8 (EMBL:AB032468) (490 aa) fasta scores: E(): 0, 39.3% id in 468 aa, and to Acinetobacter baumannii L-2,4-diaminobutyrate decarboxylase ddc SW:DDC-ACIBA (Q43908) (510 aa) fasta scores: E(): 0, 38.3% id in 431 aa" /codon_start=1 /transl_table=11 /product="putative decarboxylase" /protein_id="CAC90352.1" /db_xref="GI:15979572" /db_xref="SPTREMBL:O8ZF24" /translation="MYKDIYEENHFIDQGNIDITDVLCLFNKDSAEHYENVNNTI ARIKAVFTLVDPKPSGVLPSLSPDFYCIDLNQPLTSLDSALDELQQLYIKDAIYFHH		

```
PKYVAHLNCPVLPVAVLAESIISTNSSLDTWQSGATLIEQKVLDTWRDKMALGSGQ
GDGIFTSGGTQSNLMAILLARDHCFCHORDANHONKLOGLPADSHKFRIFTSVVSHFST
KQAAALGLGYNNAVVPYDSEKMDAQALDOSIALCLSGONIPLAIVATGTTDFGCS
IDPLHTISTIAKQYGLHWLDAAYGCGLLVSSHYEQLSGIEAMDSVTVDYHKSFQPS
VSGAFVQKQKHFYSYLYTHAEYLNPLSASLEGTPNLVNSKIQTTRFDAQKLWLTLR
IMGAQLGALVQFMDTAQAYQLMLPNPYFELTHRPVLSTLVRFRPEQKRDSDDDID
RTNAYIRKALFRECHTVIAGTKVNGKQYLFKFTFNTTTLQHLNEIVDAITEGKNYP
TLVNSQS"
1699. 2619
/misc_feature
/ gene="YPO1529"
/ note="Pfam match to entry PF00282 pyridoxal_deC,
Pyridoxal-dependent decarboxylase conserved domain, score
165.20, E-value 2.7e-47"
2885. 4183
/ gene="YPO1530"
/ note="alCB"
2885. 4183
/ gene="YPO1530"
/ note="Similar to Bordetella bronchiseptica alcaligin
biosynthesis enzyme SW:ALCA_BORBR (Q44740) (461 aa) fasta
scores: E(): 0, 49.6% id in 425 aa, and to Streptomyces
coelicolor putative monooxygenase Sccl05.14 TR:Q9L071
(EMBL:AL163641) (413 aa) fasta scores: E(): 0, 53.7% id in
419 aa"
/ codon_start=1
/ transl_table=11
/ product="putative siderophore biosynthetic enzyme"
/ protein_id="CAC90353.1"
/ db_xref="GI:15979573"
/ db_xref="SPTREMBL:Q82FZ3"
/ translation="MRPYDPIGIGIPFNGLACLTQPLNDVCLFIDQKAGFDWHPG
MLESATLPQPFMSDLVTLASPTPLSFLNYIKQGRIFYSFYIRENFPLLRKYNOC
QWAEOLPNFTQVDQDIHYDEDTALRYLCSDCTGSGRHEFLCKLVGLTGSPSHI
PECCOPELQOIVMSGQYLEHKVALQKRAITILGSGSAAEIFYDLCELDREDYQLN
WLRSRPFLETKLTLEMTSEYDVFYHLPERKDELNLTHKPIYGLFTEGAVL
IFDLYTKRGLNALNVLNTNSELQKVSSHDGKGLHLSLQHEQNLQFLETEGAVL
ATGYHQPPTFIEGIRHIRMNKGRIYEQVQVNSIDRNNTIFQVQVELTHGFTVTPDL
GMACYRNSCLIRITETGYPIEKNTAFQQSAPSLRKE"
2912. 2944
/misc_feature
/ gene="YPO1530"
/ note="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site."
4180. 4761
/ gene="YPO1531"
/ note="alCB"
4180. 4761
/ gene="YPO1531"
/ note="Similar to Bordetella bronchiseptica AlCB alcB
TR:P94254 (EMBL:U61153) (201 aa) fasta scores: E(): 2e-27,
43.7% id in 174 aa, and to Streptomyces coelicolor
putative acetyltransferase sccl05.15 TR:Q9L070
(EMBL:AL163641) (184 aa) fasta scores: E(): 3e-23, 39.5%
id in 172 aa"
/ codon_start=1
/ transl_table=11
/ product="putative siderophore biosynthetic enzyme"
/ protein_id="CAC90354.1"
/ db_xref="GI:15979574"
/ db_xref="SPTREMBL:Q82FZ2"
/ translation="MMAARNIESDTVVRSGGQLRPLVLADHVTWLTVLRNREYAHY
WGMQGVSPQVAFYQELNLSQPGVYVYQQQPAFLRLYLAQHDVADYYPADPD
DVGMHLVAPPVVERITGFTWRVFOAIMTFISDPTVARIIVPEVDVRNEKTHRLNKRAG
FVYQQQLTLPKHTAWLAFCTREQYRLALQKEHQ"
4758. 6602
/ gene="YPO1532"
4758. 6602
/ gene="YPO1532"
/ note="Similar to Bordetella bronchiseptica AlCC TR:p94255
(EMBL:U61153) (618 aa) fasta scores: E(): 0, 53.5% id in
608 aa, and to Streptomyces coelicolor hypothetical 66.6
kDa protein Sccl05.16 TR:Q9L069 (EMBL:AL163641) (595 aa)
fasta scores: E(): 0, 49.4% id in 589 aa. Similar to
YPO1538 (46.8% identity in 596 aa overlap)."
4758. 6602
```

```
/codon_start=1
/transl_table=11
/product="putative siderophore biosynthetic enzyme"
/protein_id="CAC90355.1"
/db_xref="GI:15979575"
/db_xref="SPTREMBL:Q82FZ1"
/translation="MMDLASPLSPHQSLELPHLOPVLMQKNRNLHCKAISFSHEC
LLAPQRTDHPDSEGYDYQLVAAADKXANVYFRRLALDHLWDPSLHKTVNEK
STDLDLLFIIEFKRQISISERVLYLEITTSYSSAFKHCRTGISATLVNASFQ
IIEKMEHGHSFVANNRIGEDAQDFQRPSEASDVHLVWLAAHKSFAHACIQL
DYAKLMEQELGAELAEFEQQLIARDQNFQVPMFVHPWQMNKLTISFAADIANQR
LVFLGGQKDAYQAQQSIRTFNFRSHPRQYVVKMALSIILNMGFMKGLSPYMMATPCIN
EWLFDVEGDEILOAYDFKILREVASIGFRNSYVEQAITGDSAYKKMAAALNRENPLQ
LIOPGQOLMTMALHLHQEGKALLPALIDASGLSTQOHLERYLSRYSPLHLCLLYV
DLMPHFGENIILVLQNAVPOHIFMKDLAEELVLVNTADLPKAKRIIVDMPEMKT
LTLSDVDFGVFRYLAADVDDQGDYPERYFADVAACIRYOQQYPELAAKFAFYPLF
TPKIKRCLNRLQNLNNHMLDLADVPQSFQFADDLINPLVEFEVSAG"
complement(6711. 7559)
/ gene="YPO1533"
complement(6711. 7559)
/ gene="YPO1533"
/ note="Similar to Vibrio cholerae ferric vibriobactin ABC
transporter, ATP-binding protein vc0779 TR:Q9KTV8
(EMBL:AE004163) (279 aa) fasta scores: E(): 0, 55.9% id in
263 aa, and to Streptomyces coelicolor putative
iron-siderophore uptake system ATP-binding component
sc151.25C TR:Q9S215 (EMBL:AL109848) (301 aa) fasta scores:
E(): 0, 57.2% id in 264 aa"
/ codon_start=1
/ transl_table=11
/ product="putative iron-siderophore transport system,
ATP-binding component"
/ protein_id="CAC90356.1"
/ db_xref="GI:15979576"
/ db_xref="SPTREMBL:Q82FZ0"
/ translation="MRSQRELTSPMTPLENPGDISDKPRLTTQKLSLGYGQOLINKL
SLIKNGAFSVIIGPNCSTLLRALSLLPONGSIRLDQODIOHYKAKVFAROLS
LLSOGASISETVFDLVSRYGHOSFFHOMSTEDERIVKKAISLVNLSLVQQRVS
ELSGGORAFAMFAMLAOHTPTVLLDEPTVLDIAHQEMLDLQBLQCGKTLVLVL
HDINQALRATHLIMLKEQIYAEGLPESIVTDETSAAVFLGKRIITDPESGKPLVI
PRQSR"
complement(6849. 7409)
/ gene="YPO1533"
/ note="Pfam match to entry PF00005 ABC_tran, ABC
transporter, score 183.50, E-value 3.5e-51"
complement(7032. 7076)
/ gene="YPO1533"
/ note="PS00211 ABC transporters family signature."
complement(7365. 7388)
/ gene="YPO1533"
/ note="PS00017 ATP/GTP-binding site motif A (P-loop)."
complement(7556. 8596)
/ gene="YPO1534"

Query Match 3.0%; Score 20; DB 1; Length 193050;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 520 CTTTATGAAGCGCGCAACCG 539
|||||
Db 127609 CTTTATGAAGCGCGCAACCG 127590

RESULT 20
AL353689/c
LOCUS
DEFINITION
AL353689
ACCESSION
AL353689
VERSION
AL353689.26
KEYWORDS
HTG.
SOURCE
human.
ORGANISM
Homo sapiens
```


Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 195829)
Moore,W.
Direct Submission
Submitted (19-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Oct 21, 2001 this sequence version replaced gi:15983876.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chrl
RP5-1087E8 is from the library RPCI-5 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: PCYPAC2

This sequence is the entire insert of clone RP5-1087E8 The true
left end of clone RP11-1B20 is at 155425 in this sequence. The true
right end of clone RP11-375H24 is at 75234 in this sequence.

FEATURES
source
1..195829
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="q42.11-42.2"
/clone="RP5-1087E8"
/clone_lib="RPCI-5"
13995..14136
/note="Sequence from overlapping clone RP11-375H24
(AL391628). Assembly confirmed by restriction digest."
17759
/note="Tandem repeat. Forced join. Gap size estimated to
be approximately 250bp by restriction digest data."
22581
/note="Tandem repeat. Forced join. Gap size estimated to
be approximately 600bps by restriction digest data."
38361
complement(38872..39056)
/note="Sequence from overlapping clone RP11-375H24
(AL391628). Assembly confirmed by restriction digest."
39840..39867
/note="Single clone region. Assembly confirmed by
restriction digest data."
195130..195532
/note="Sequence from overlapping clone RP11-1B20
(AL451047). Assembly confirmed by restriction digest."
BASE COUNT 57449 a 40631 c 39653 g 58096 t
ORIGIN

Query Match 3.0%; Score 20; DB 9; Length 195829;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 AAAATTTACTATGTCAGGT 344

Db 150492 AAAATTTACTATGTCAGGT 150473
|||||
Search completed: January 21, 2003, 00:05:22
Job time : 3147 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: January 20, 2003, 23:02:09 ; Search time 230 Seconds
(without alignments)
6550.378 Million cell updates/sec

Title: US-09-252-691C-1394

Perfect score: 669

Sequence: 1 gctataatggcgcaatgat.....ttacccttaaggagaactga 669

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 20

Total number of hits satisfying chosen parameters: 3

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : N_Geneseq_101002.*

1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	23	3.4	1881	23 AAS85924	DNA encoding novel
C 2	20	3.0	522	24 ABQ91613	M. capsulatus gene
C 3	20	3.0	603	24 ABQ91612	M. capsulatus gene

ALIGNMENTS

RESULT 1

AAS85924/c
ID AAS85924 standard; cDNA; 1881 BP.
XX
AC AAS85924;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #21728.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG21737.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 1; SEQ ID No 21728; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1881 BP; 419 A; 469 C; 533 G; 460 T; 0 other;
Query Match 3.4%; Score 23; DB 23; Length 1881;
Best Local Similarity 100.0%; Pred. No. 0.074;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 133 CTGTTCAATAAACCCCTACCATGT 155
|||||
Db 591 CTGTTCAATAAACCCCTACCATGT 569
RESULT 2
ABQ91613
ID ABQ91613 standard; DNA; 522 BP.

```
XX ABQ91613;
AC
XX
XX 01-OCT-2002 (first entry)
DT
XX
XX M. capsulatus gene #1598 for DNA array.
DE
XX
XX Micro array; gene; ds; differential expression; gene expression.
KW
XX
XX Methylococcus capsulatus.
OS
XX
XX WO200255655-A2.
PN
XX
XX 18-JUL-2002.
PD
XX
XX 14-JAN-2002; 2002WO-NO00019.
PF
XX
XX 12-JAN-2001; 2001NO-0000235.
PR
XX
XX 12-JAN-2001; 2001NO-0000239.
PR
XX
XX (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.
PA
XX (TIGR-) TIGR.
XX
XX Birkeland NK, Eidhammer I, Jonassen I, Jensen HB, Lien T;
PI Lillehaug JR, Lossius I, Eisen JA, Fraser CM, Durkin AS;
PI Salzberg SL;
XX
XX WPI; 2002-557818/59.
DR
XX
XX Novel DNA array useful for determining differential expression of
XX Methylococcus capsulatus genes, comprises polynucleotides or
XX oligonucleotides representative for a selective number of Methylococcus
XX capsulatus genes -
XX
XX Claim 14; Page 609; 678pp; English.
PS
XX
XX The invention relates to a novel DNA array giving a representation of a
XX number of Methylococcus capsulatus genes. The method of the invention is
XX useful for determination of the differential expression of the genes of
XX M. capsulatus, and for studying gene expression on a genomic scale and in
XX gene expression assays of M. capsulatus genes. The sequences shown in
XX ABQ90016-ABQ91855 represent M. capsulatus genes for use in arrays of the
XX invention.
XX
XX Sequence 522 BP; 107 A; 152 C; 145 G; 118 T; 0 other;
SQ
Query Match 3.0%; Score 20; DB 24; Length 522;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 245 GCGACAGCGAGGGGCTGCTG 264
Db 383 GCGACAGCGAGGGGCTGCTG 402
|||||
RESULT 3
ABQ91612/c
ID ABQ91612 standard; DNA; 603 BP.
XX
XX
AC ABQ91612;
XX
XX 01-OCT-2002 (first entry)
DT
XX
XX M. capsulatus gene #1597 for DNA array.
DE
XX
XX Micro array; gene; ds; differential expression; gene expression.
KW
XX
XX Methylococcus capsulatus.
OS
XX
XX WO200255655-A2.
PN
XX
XX 18-JUL-2002.
PD
XX
XX
```

```
PF 14-JAN-2002; 2002WO-NO00019.
XX
XX 12-JAN-2001; 2001NO-0000235.
PR
XX 12-JAN-2001; 2001NO-0000239.
XX
XX (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.
PA (TIGR-) TIGR.
XX
XX Birkeland NK, Eidhammer I, Jonassen I, Jensen HB, Lien T;
PI Lillehaug JR, Lossius I, Eisen JA, Fraser CM, Durkin AS;
PI Salzberg SL;
XX
XX WPI; 2002-557818/59.
DR
XX
XX Novel DNA array useful for determining differential expression of
XX Methylococcus capsulatus genes, comprises polynucleotides or
XX oligonucleotides representative for a selective number of Methylococcus
XX capsulatus genes -
XX
XX Claim 14; Page 608-609; 678pp; English.
PS
XX
XX The invention relates to a novel DNA array giving a representation of a
XX number of Methylococcus capsulatus genes. The method of the invention is
XX useful for determination of the differential expression of the genes of
XX M. capsulatus, and for studying gene expression on a genomic scale and in
XX gene expression assays of M. capsulatus genes. The sequences shown in
XX ABQ90016-ABQ91855 represent M. capsulatus genes for use in arrays of the
XX invention.
XX
XX Sequence 603 BP; 135 A; 169 C; 180 G; 119 T; 0 other;
SQ
Query Match 3.0%; Score 20; DB 24; Length 603;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 245 GCGACAGCGAGGGGCTGCTG 264
Db 233 GCGACAGCGAGGGGCTGCTG 214
|||||
Search completed: January 20, 2003, 23:12:43
Job time : 231 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2003, 23:07:39 ; Search time 50 Seconds
(without alignments)
4103.335 Million cell updates/sec

Title: US-09-252-691C-1394
Perfect score: 669
Sequence: 1 gctataatggcgcaattgat.....ttacccttaaggagaactga 669

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 20

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES			
Result	Query		
No.	Score	Match Length DB ID	Description

No matches found

Search completed: January 21, 2003, 00:30:41
Job time : 50 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2003, 23:01:24 ; Search time 56 Seconds
(without alignments)
5326.534 Million cell updates/sec

Title: US-09-252-691C-1394
Perfect score: 669
Sequence: 1 gctataatgcggcaatgat.....ttacccttaaggagaactga 669

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 393868 seqs, 222934149 residues

Word size : 20

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Listing first 1000 summaries

- Database : Published_Applications_NA:*
- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
 - 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
 - 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
 - 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
 - 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
 - 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
 - 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
 - 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
 - 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
 - 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
 - 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
 - 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
 - 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
 - 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

No matches found

Search completed: January 20, 2003, 23:08:37
Job time : 56 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2003, 23:07:04 ; Search time 1959 Seconds
(without alignments)
5530.771 Million cell updates/sec

Title: US-09-252-691C-1394
Perfect score: 669
Sequence: 1 gctataatggcgcaatgat.....ttaccacctagaagaactga 669

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0
Searched: 16154066 seqs, 8097743376 residues
Word size : 20

Total number of hits satisfying chosen parameters: 7

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	22	3.3	497	17	AZ296025
2	20	3.0	522	10	BE115864
3	20	3.0	565	10	AV523923
4	20	3.0	699	17	BH160291
5	20	3.0	818	17	BH570517
6	20	3.0	964	17	CNS06YXS

c 7 20 3.0 1075 13 BR249598 BR249598 602996256

ALIGNMENTS

RESULT 1
AZ296025
LOCUS
DEFINITION
PCIC-23-62013.TJB RPCI-23 Mus musculus genomic clone RPCI-23-62013,
DNA sequence.
ACCESSION
AZ296025
VERSION
AZ296025.1 GI:9537810
KEYWORDS
GSS.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 497)
REFERENCE
Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
TITLE
Mouse BAC End Sequences from Library RPCI-23
JOURNAL
Unpublished (1999)
COMMENT
Other GSSs: RPCI-23-62013.TV RPCI-23-62013.TVB
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 62 row: 0 column: 13
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1. 497
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-62013"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
ECORI; Site_2: ECORI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of ECORI and ECORI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
ECORI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life technologies)."
BASE COUNT 134 a 123 c 74 g 166 t
ORIGIN

FEATURES
source

Query Match 3.3%; Score 22; DB 17; Length 497;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 44 AAACCTCTTTTAAACACCG 65
Db 24 AAACCTCTTTTAAACACCG 45
RESULT 2
BE115864/c
LOCUS
DEFINITION
UI-R-BS1-axx-c-03-0-UI.s1 UI-R-BS1 Rattus norvegicus cDNA clone

```

ACCESSION      BE115864
VERSION        BE115864.1  GI:8508009
KEYWORDS
SOURCE         Norway rat
ORGANISM       Rattus norvegicus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
               Rattus.
REFERENCE      1 (bases 1 to 522)
AUTHORS       Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE         Normalization and subtraction: two approaches to facilitate gene
               discovery
JOURNAL       Genome Res. 6 (9), 791-806 (1996)
MEDLINE       9704477
COMMENT       Contact: Soares, MB
               Program for Rat Gene Discovery and Mapping
               University of Iowa
               451 Eckstein Medical Research Building Iowa City, IA 52242, USA
               Tel: 319 335 8250
               Fax: 319 335 9565
               Email: msoares@blue.weeg.uiowa.edu
               The sequence contained an oligo-dT track that was present in the
               oligonucleotide that was used to prime the synthesis of first
               strand cDNA and therefore this may represent a bonafide poly A
               tail. The sequence tag present in the cDNA between the NotI site
               and the oligo-dT track served to verify it as a clone from the
               normalized embryo at 13 dpc library cDNA Library Preparation: M.B.
               Soares Lab Clone distribution: clones will be available through
               Research Genetics (www.resgen.com)
               Seq primer: M13 Forward
               POLYA=yes.
FEATURES       Location/Qualifiers
               1..522
               /organism="Rattus norvegicus"
               /strain="Sprague-Dawley"
               /db_xref="taxon:10116"
               /clone="UI-R-BS1-axx-c-03-0-UI"
               /clone_lib="UI-R-BS1"
               /dev_stage="embryonic 13 dpc"
               /lab_host="DH10B (Life Technologies)"
               /note="vector: pT7T3D-Pac (Pharmacia) with a modified
               polylinker; site_1: Not I; Site_2: Eco RI; The UI-R-BS1
               library is derived from 13 dpc whole embryo tissue. For a
               detailed description of the library from which this clone
               was derived, please visit our web site at
               ratest.eng.uiowa.edu.
               TAG_LIB=UI-R-BS1
               TAG_TISSUE=embryo at 13 dpc
               TAG_SEQ=ATCC"
BASE COUNT    141 a 106 c 126 g 149 t
ORIGIN
Query Match      3.0%; Score 20; DB 10; Length 522;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 537 CCGACAGTCAGCGGCATGA 556
      |||||||||||||||||||
Db 298 CCGACAGTCAGCGGCATGA 279

RESULT 3
LOCUS      AV523923      565 bp  mRNA  linear  EST 01-SEP-2000
DEFINITION AV523923 Arabidopsis thaliana aboveground organs two to six-week
               old Arabidopsis thaliana cDNA clone AP2L44a05F 3', mRNA sequence.
ACCESSION  AV523923
VERSION    AV523923
KEYWORDS   AV523923.1 GI:8603451
SOURCE     Est.
ORGANISM   Arabidopsis thaliana
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

UI-R-BS1-axx-c-03-0-UI 3', mRNA sequence.
BE115864
BE115864.1  GI:8508009
EST.
Norway rat
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 522)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
9704477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
normalized embryo at 13 dpc library cDNA Library Preparation: M.B.
Soares Lab Clone distribution: clones will be available through
Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=yes.
FEATURES       Location/Qualifiers
               1..522
               /organism="Rattus norvegicus"
               /strain="Sprague-Dawley"
               /db_xref="taxon:10116"
               /clone="UI-R-BS1-axx-c-03-0-UI"
               /clone_lib="UI-R-BS1"
               /dev_stage="embryonic 13 dpc"
               /lab_host="DH10B (Life Technologies)"
               /note="vector: pT7T3D-Pac (Pharmacia) with a modified
               polylinker; site_1: Not I; Site_2: Eco RI; The UI-R-BS1
               library is derived from 13 dpc whole embryo tissue. For a
               detailed description of the library from which this clone
               was derived, please visit our web site at
               ratest.eng.uiowa.edu.
               TAG_LIB=UI-R-BS1
               TAG_TISSUE=embryo at 13 dpc
               TAG_SEQ=ATCC"
BASE COUNT    141 a 106 c 126 g 149 t
ORIGIN
Query Match      3.0%; Score 20; DB 10; Length 522;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 537 CCGACAGTCAGCGGCATGA 556
      |||||||||||||||||||
Db 298 CCGACAGTCAGCGGCATGA 279

RESULT 3
LOCUS      AV523923      565 bp  mRNA  linear  EST 01-SEP-2000
DEFINITION AV523923 Arabidopsis thaliana aboveground organs two to six-week
               old Arabidopsis thaliana cDNA clone AP2L44a05F 3', mRNA sequence.
ACCESSION  AV523923
VERSION    AV523923
KEYWORDS   AV523923.1 GI:8603451
SOURCE     Est.
ORGANISM   Arabidopsis thaliana
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 565)
Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
20363093
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizuekazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES       Location/Qualifiers
               1..565
               /organism="Arabidopsis thaliana"
               /strain="Columbia"
               /db_xref="taxon:3702"
               /clone="AP2L44a05F"
               /clone_lib="Arabidopsis thaliana aboveground organs two to
               six-week old"
               /tissue_type="aboveground organs"
               /dev_stage="two to six-week old"
               /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
               XhoI"
BASE COUNT    153 a 125 c 128 g 159 t
ORIGIN
Query Match      3.0%; Score 20; DB 10; Length 565;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 588 CATTCGCTAGCCATGGCA 607
      |||||||||||||||||||
Db 543 CATTCGCTAGCCATGGCA 562

RESULT 4
LOCUS      BH160291/c      699 bp  DNA  linear  GSS 24-SEP-2001
DEFINITION BH160291 Entamoeba histolytica Sheared DNA Entamoeba histolytica
               genomic, DNA sequence.
ACCESSION  BH160291
VERSION    BH160291.1  GI:15733729
KEYWORDS   GSS.
SOURCE     Entamoeba histolytica.
ORGANISM   Entamoeba histolytica
               Eukaryota; Entamoebidae; Entamoeba.
REFERENCE  1 (bases 1 to 699)
AUTHORS    Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE      Determination of clone end sequences from Entamoeba histolytica
               Hm1:IMSS sheared DNA library (2001)
JOURNAL    Unpublished (2001)
COMMENT    Contact: Brendan J Loftus
               Department of Eukaryotic Genomics
               The Institute for Genomic Research
               9712 Medical Center Dr., Rockville, MD 20850, USA
               Tel: 301 838 0208
               Fax: 301 838 3543
               Email: bjlloftus@tigr.org
               Clones are derived from the Entamoeba histolytica Hm1:IMSS sheared
               DNA library
               Seq primer: M13-Reverse
               Class: shotgun
               High quality sequence start: 10
               High quality sequence stop: 264.
               Location/Qualifiers
               1..699
               /organism="Entamoeba histolytica"
               /strain="Hm1:IMSS"
               /db_xref="taxon:5759"
               /clone_lib="Entamoeba histolytica Sheared DNA"

```


4